

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : A61K 37/10, 37/36	A1	(11) International Publication Number: WO 94/26298 (43) International Publication Date: 24 November 1994 (24.11.94)
(21) International Application Number: PCT/US94/05083 (22) International Filing Date: 6 May 1994 (06.05.94) (30) Priority Data: 08/059,022 6 May 1993 (06.05.93) US 08/209,204 8 March 1994 (08.03.94) US (71) Applicant (for all designated States except US): CAMBRIDGE NEUROSCIENCE [US/US]; Building 700, One Kendall Square, Cambridge, MA 02139 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): SKLAR, Robert [US/US]; 36 Spiers Road, Newton, MA 02159 (US). MARCHIONNI, Mark [US/US]; 24 Twin Circle Drive, Arlington, MA 02174 (US). GWYNNE, David, I. [GB/US]; 77 Grover Street, Beverly, MA 01915 (US). (74) Agent: CLARK, Paul, T.; Fish & Richardson, 225 Franklin Street, Boston, MA 02110 (US).	(81) Designated States: AU, CA, JP, KR, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(54) Title: METHODS FOR TREATING MUSCLE DISEASES AND DISORDERS (57) Abstract The invention relates to methods of treating diseases and disorders of the muscle tissues in a vertebrate by the administration of compounds which bind the p185 ^{erbB2} receptor. These compounds are found to cause increased differentiation and survival of cardiac, skeletal and smooth muscle.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

METHODS FOR TREATING MUSCLE DISEASES AND DISORDERS

Background of the Invention

The invention relates to prophylactic or affirmative
5 treatment of diseases and disorders of the musculature by
administering polypeptides found in vertebrate species,
which polypeptides are growth, differentiation and survival
factors for muscle cells.

Muscle tissue in adult vertebrates will regenerate
10 from reserve myoblasts called satellite cells. Satellite
cells are distributed throughout muscle tissue and are
mitotically quiescent in the absence of injury or disease.
Following muscle injury or during recovery from disease,
satellite cells will reenter the cell cycle, proliferate and
15 1) enter existing muscle fibers or 2) undergo
differentiation into multinucleate myotubes which form new
muscle fiber. The myoblasts ultimately yield replacement
muscle fibers or fuse into existing muscle fibers, thereby
increasing fiber girth by the synthesis of contractile
20 apparatus components. This process is illustrated, for
example, by the nearly complete regeneration which occurs in
mammals following induced muscle fiber degeneration; the
muscle progenitor cells proliferate and fuse together
regenerating muscle fibers.

25 Several growth factors which regulate the
proliferation and differentiation of adult (and embryonic)
myoblasts in vitro have been identified. Fibroblast growth
factor (FGF) is mitogenic for muscle cells and is an
inhibitor of muscle differentiation. Transforming growth
30 factor β (TGF β) has no effect on myoblast proliferation, but
is an inhibitor of muscle differentiation. Insulin-like
growth factors (IGFs) have been shown to stimulate both
myoblast proliferation and differentiation in rodents.
Platelet derived growth factor (PDGF) is also mitogenic for
35 myoblasts and is a potent inhibitor of muscle cell

- 2 -

differentiation see: Florini and Magri, 1989:256:C701-C711).

In vertebrate species both muscle tissue and neurons are potential sources of factors which stimulate myoblast proliferation and differentiation. In diseases affecting the neuromuscular system which are neural in origin (i.e., neurogenic), the muscle tissue innervated by the affected nerve becomes paralyzed and wastes progressively. During peripheral nerve regeneration and recovery from neurologic and myopathic disease, neurons may provide a source of growth factors which elicit the muscle regeneration described above and provide a mechanism for muscle recovery from wasting and atrophy.

A recently described family of growth factors, the neuregulins, are synthesized by motor neurons (Marchionni et al. *Nature* 362:313, 1993) and inflammatory cells (Tarakhovsky et al., *Oncogene* 6:2187-2196 (1991)). The neuregulins and related p185^{erbB2} binding factors have been purified, cloned and expressed (Benveniste et al., *PNAS* 82:3930-3934, 1985; Kimura et al., *Nature* 348:257-260, 1990; Davis and Stroobant, *J. Cell. Biol.* 110:1353-1360, 1990; Wen et al., *Cell* 69:559, 1992; Yarden and Ullrich, *Ann. Rev. Biochem.* 57:443, 1988; Holmes et al., *Science* 256:1205, 1992; Dobashi et al., *Proc. Natl. Acad. Sci.* 88:8582, 1991; Lupu et al., *Proc. Natl. Acad. Sci.* 89:2287, 1992). Recombinant neuregulins have been shown to be mitogenic for peripheral glia (Marchionni et al., *Nature* 362:313, 1993) and have been shown to influence the formation of the neuromuscular junction (Falls et al., *Cell* 72:801, 1993). Thus the regenerating neuron and the inflammatory cells associated with the recovery from neurogenic disease and nerve injury provide a source of factors which coordinate the remyelination of motor neurons and their ability to form the appropriate connection with their target. After

- 3 -

muscle has been reinnervated the motor neuron may provide factors to muscle, stimulating muscle growth and survival.

Currently, there is no useful therapy for the
5 promotion of muscle differentiation and survival. Such a therapy would be useful for treatment of a variety of neural and muscular diseases and disorders.

Summary of the Invention

We have discovered that increased mitogenesis
10 differentiation and survival of muscle cells may be achieved using proteins heretofore described as glial growth factors, acetylcholine receptor inducing activity (ARIA), heregulins, neu differentiation factor, and, more generally, neuregulins. We have discovered that these
15 compounds are capable of inducing both the proliferation of muscle cells and the differentiation and survival of myotubes. These phenomena may occur in cardiac and smooth muscle tissues in addition to skeletal muscle tissues. Thus, the above compounds, regulatory compounds
20 which induce synthesis of these compounds, and small molecules which mimic these compounds by binding to the receptors on muscle or by stimulating through other means the second messenger systems activated by the ligand-receptor complex are all extremely useful as prophylactic
25 and affirmative therapies for muscle diseases.

A novel aspect of the invention involves the use of the above named proteins as growth factors to induce the mitogenesis, survival, growth and differentiation of muscle cells. Treating of the muscle cells to achieve
30 these effects may be achieved by contacting muscle cells with a polypeptide described herein. The treatments may be provided to slow or halt net muscle loss or to

- 4 -

increase the amount or quality of muscle present in the vertebrate.

These factors may be used to produce muscle cell mitogenesis, differentiation, and survival in a vertebrate (preferably a mammal, more preferably a human) by administering to the vertebrate an effective amount of a polypeptide or a related compound. Neuregulin effects on muscle may occur, for example, by causing an increase in muscle performance by inducing the synthesis of particular isoforms of the contractile apparatus such as the myosin heavy chain slow and fast isoforms; by promoting muscle fiber survival via the induction of synthesis of protective molecules such as, but not limited to, dystrophin; and/or by increasing muscle innervation by, for example, increasing acetylcholine receptor molecules at the neuromuscular junction.

The term muscle cell as used herein refers to any cell which contributes to muscle tissue. Myoblasts, satellite cells, myotubes, and myofibril tissues are all included in the term "muscle cells" and may all be treated using the methods of the invention. Muscle cell effects may be induced within skeletal, cardiac and smooth muscles.

Mitogenesis may be induced in muscle cells, including myoblasts or satellite cells, of skeletal muscle, smooth muscle or cardiac muscle. Mitogenesis as used herein refers to any cell division which results in the production of new muscle cells in the patient. More specifically, mitogenesis in vitro is defined as an increase in mitotic index relative to untreated cells of 50%, more preferably 100%, and most preferably 300%, when the cells are exposed to labelling agent for a time equivalent to two doubling times. The mitotic index is the fraction of cells in the culture which have labelled

- 5 -

nuclei when grown in the presence of a tracer which only incorporates during S phase (i.e., BrdU) and the doubling time is defined as the average time required for the number of cells in the culture to increase by a factor of two).

An effect on mitogenesis in vivo is defined as an increase in satellite cell activation as measured by the appearance of labelled satellite cells in the muscle tissue of a mammal exposed to a tracer which only incorporates during S phase (i.e., BrdU). In useful therapeutic is defined in vivo as a compound which increases satellite cell activation relative to a control mammal by at least 10%, more preferably by at least 50%, and most preferably by more than 200% when the mammal is exposed to labelling agent for a period of greater than 15 minutes and tissues are assayed between 10 hours and 24 hours after administration of the mitogen at the therapeutic dose. Alternatively, satellite cell activation in vivo may be detected by monitoring the appearance of the intermediate filament vimentin by immunological or RNA analysis methods. When vimentin is assayed, the useful mitogen is defined as one which causes expression of detectable levels of vimentin in the muscle tissue when the therapeutically useful dosage is provided.

Myogenesis as used herein refers to any fusion of myoblasts to yield myotubes. Most preferably, an effect on myogenesis is defined as an increase in the fusion of myoblasts and the enablement of the muscle differentiation program. The useful myogenic therapeutic is defined as a compound which confers any increase in the fusion index in vitro. More preferably, the compound confers at least a 2.0-fold increase and, most preferably, the compound confers a 3-fold or greater

- 6 -

increase in the fusion index relative to the control. The fusion index is defined as the fraction of nuclei present in multinucleated cells in the culture relative to the total number of nuclei present in the culture.

- 5 The percentages provided above are for cells assayed after 6 days of exposure to the myogenic compound and are relative to an untreated control. Myogenesis may also be determined by assaying the number of nuclei per area in myotubes or by measurement of the levels of muscle
10 specific protein by Western analysis. Preferably, the compound confers at least a 2.0-fold increase in the density of myotubes using the assay provided, for example, herein, and, most preferably, the compound confers a 3-fold or greater increase.

- 15 The growth of muscle may occur by the increase in the fiber size and/or by increasing the number of fibers. The growth of muscle as used herein may be measured by A) an increase in wet weight, B) an increase in protein content, C) an increase in the number of muscle fibers,
20 or D) an increase in muscle fiber diameter. An increase in growth of a muscle fiber can be defined as an increase in the diameter where the diameter is defined as the minor axis of ellipsis of the cross section. The useful therapeutic is one which increases the wet weight,
25 protein content and/or diameter by 10% or more, more preferably by more than 50% and most preferably by more than 100% in an animal whose muscles have been previously degenerated by at least 10% and relative to a similarly treated control animal (i.e., an animal with degenerated
30 muscle tissue which is not treated with the muscle growth compound). A compound which increases growth by increasing the number of muscle fibers is useful as a therapeutic when it increases the number of fibers in the diseased tissue by at least 1%, more preferably at least

- 7 -

20%, and most preferably, by at least 50%. These percentages are determined relative to the basal level in a comparable untreated undiseased mammal or in the contralateral undiseased muscle when the compound is administered and acts locally.

The survival of muscle fibers as used herein refers to the prevention of loss of muscle fibers as evidenced by necrosis or apoptosis or the prevention of other mechanisms of muscle fiber loss. Survival as used herein indicates an decrease in the rate of cell death of at least 10%, more preferably by at least 50%, and most preferably by at least 300% relative to an untreated control. The rate of survival may be measured by counting cells stainable with a dye specific for dead cells (such as propidium iodide) in culture when the cells are 8 days post-differentiation (i.e., 8 days after the media is changed from 20% to 0.5% serum).

Muscle regeneration as used herein refers to the process by which new muscle fibers form from muscle progenitor cells. The useful therapeutic for regeneration confers an increase in the number of new fibers by at least 1%, more preferably by at least 20%, and most preferably by at least 50%, as defined above.

The differentiation of muscle cells as used herein refers to the induction of a muscle developmental program which specifies the components of the muscle fiber such as the contractile apparatus (the myofibril). The therapeutic useful for differentiation increases the quantity of any component of the muscle fiber in the diseased tissue by at least 10% or more, more preferably by 50% or more, and most preferably by more than 100% relative to the equivalent tissue in a similarly treated control animal.

- 8 -

Atrophy of muscle as used herein refers to a significant loss in muscle fiber girth. By significant atrophy is meant a reduction of muscle fiber diameter in diseased, injured or unused muscle tissue of at least 10% relative to undiseased, uninjured, or normally utilized tissue.

Methods for treatment of diseases or disorders using the polypeptides or other compounds described herein are also part of the invention. Examples of muscular disorders which may be treated include skeletal muscle diseases and disorders such as myopathies, dystrophies, myoneural conductive diseases, traumatic muscle injury, and nerve injury. Cardiac muscle pathologies such as cardiomyopathies, ischemic damage, congenital disease, and traumatic injury may also be treated using the methods of the invention, as may smooth muscle diseases and disorders such as arterial sclerosis, vascular lesions, and congenital vascular diseases. For example, Duchennes muscular dystrophy, Beckers' dystrophy, and Myasthenia gravis are but three of the diseases which may be treated using the methods of the invention.

The invention also includes methods for the prophylaxis or treatment of a tumor of muscle cell origin such as rhabdomyosarcoma. These methods include administration of an effective amount of a substance which inhibits the binding of one or more of the polypeptides described herein and inhibiting the proliferation of the cells which contribute to the tumor.

The methods of the invention may also be used to treat a patient suffering from a disease caused by a lack of a neurotrophic factor. By lacking a neurotrophic factor is meant a decreased amount of neurotrophic factor relative to an unaffected individual sufficient to cause

- 9 -

detectable decrease in neuromuscular connections and/or muscular strength. The neurotrophic factor may be present at levels 10% below those observed in unaffected individuals. More preferably, the factor is present at
5 levels 20% lower than are observed in unaffected individuals, and most preferably the levels are lowered by 80% relative to unaffected individuals under similar circumstances.

The methods of the invention make use of the fact
10 that the neuregulin proteins are encoded by the same gene. A variety of messenger RNA splicing variants (and their resultant proteins) are derived from this gene and many of these products show binding to P185^{erbB2} and activation of the same. Products of this gene have been
15 used to show muscle cell mitogenic activity (see Examples 1 and 2, below), differentiation (Examples 3 and 6), and survival (Examples 4 and 5). This invention provides a use for all of the known products of the neuregulin gene (described herein and in the references listed above)
20 which have the stated activities as muscle cell mitogens, differentiation factors, and survival factors. Most preferably, recombinant human GGF2 (rhGGF2) is used in these methods.

The invention also relates to the use of other,
25 not yet naturally isolated, splicing variants of the neuregulin gene. Fig. 29 shows the known patterns of splicing. These patterns are derived from polymerase chain reaction experiments (on reverse transcribed RNA), analysis of cDNA clones (as presented within), and
30 analysis of published sequences encoding neuregulins (Peles et al., Cell 69:205 (1992) and Wen et al., Cell 69:559 (1992)). These patterns, as well as additional patterns disclosed herein, represent probable splicing variants which exist. The splicing variants are fully

- 10 -

described in Goodearl et al., USSN 08/036,555, filed March 24, 1993, incorporated herein by reference.

More specifically, cell division, survival, differentiation and growth of muscle cells may be
 5 achieved by contacting muscle cells with a polypeptide defined by the formula

WBAZCX

wherein WBAZCX is composed of the polypeptide segments shown in Fig. 30 (SEQ ID Nos. 132, 134, 135,
 10 137-139, 156); wherein W comprises the polypeptide segment F, or is absent; wherein Z comprises the polypeptide segment G or is absent; and wherein X comprises the polypeptide segment C/D HKL, C/D H, C/D HL, C/D D, C/D' HL, C/D' HKL, C/D' H, C/D' D, C/D C/D' HKL,
 15 C/D C/D' H, C/D C/D' HL, C/D C/D' D, C/D D' H, C/D D' HL, C/D D' HKL, C/D' D' H, C/D' D' HL, C/D' D' HKL, C/D C/D' D' H, C/D C/D' D' HL, or C/D C/D' D' HKL and/or by contacting muscle cells with a polypeptide defined by the formula

20

YBAZCX

wherein YBAZCX is composed of the polypeptide segments shown in Fig. 30 (SEQ ID Nos. 133-135, 156, 159); wherein Y comprises the polypeptide segment E, or is absent; wherein Z comprises the polypeptide segment G
 25 or is absent; and wherein X comprises the polypeptide segment C/D HKL, C/D H, C/D HL, C/D D, C/D' HL, C/D' HKL, C/D' H, C/D' D, C/D C/D' HKL, C/D C/D' H, C/D C/D' HL, C/D C/D' D, C/D D' H, C/D D' HL, C/D D' HKL, C/D' D' H, C/D' D' HL, C/D' D' HKL, C/D C/D' D' H, C/D C/D' D' HL,
 30 or C/D C/D' D' HKL.

Generally, the N-terminus of the above-described polypeptides begins with either the F or E polypeptide

- 11 -

segments. When the F polypeptide is present it may be cleaved upon maturation of the protein to yield the mature polypeptide. When the E sequence is present the first 50 amino acids which represent the N-terminal
5 signal sequence may be absent from the polypeptides.

Furthermore, the invention includes a method of treating muscle cells by the application to the muscle cell of a

- 30 kD polypeptide factor isolated from the MDA-MB
10 231 human breast cell line; or
- 35 kD polypeptide factor isolated from the rat I-EJ transformed fibroblast cell line to the glial cell or
- 75 kD polypeptide factor isolated from the SKBR-3
human breast cell line; or
- 15 -44 kD polypeptide factor isolated from the rat I-EJ transformed fibroblast cell line; or
- 25 kD polypeptide factor isolated from activated mouse peritoneal macrophages; or
- 45 kD polypeptide factor isolated from the MDA -
20 MB 231 human breast cell; or
- 7 to 14 kD polypeptide factor isolated from the ATL-2 human T-cell line to the glial cell; or
- 25 kD polypeptide factor isolated from the bovine kidney cells; or
- 25 -42 kD ARIA polypeptide factor isolated from brain; -46-47 kD polypeptide factor which stimulates O-2A glial progenitor cells; or
- 43-45 kD polypeptide factor, GGFIII, 175
U.S. patent application Serial No. 07/931,041, filed
30 August 17, 1992, incorporated herein by reference.

The invention further includes methods for the use of the EGFL1, EGFL2, EGFL3, EGFL4, EGFL5, and EGFL6 polypeptides, Fig. 37 to 42 and SEQ ID Nos. 150 to 155,

- 12 -

respectively, for the treatment of muscle cells in vivo and in vitro.

Also included in the invention is the administration of the GGF2 polypeptide whose sequence is
5 shown in Fig. 44 for the treatment of muscle cells.

An additional important aspect of the invention are methods for treating muscle cells using:

(a) a basic polypeptide factor also known to have glial cell mitogenic activity, in the presence of fetal
10 calf plasma, a molecular weight of from about 30 kD to about 36 kD, and including within its amino acid sequence any one or more of the following peptide sequences:

	F K G D A H T E	(SEQ ID NO: 1)
	A S L A D E Y E Y M X K	(SEQ ID NO: 2)
15	T E T S S S G L X L K	(SEQ ID NO: 3)
	A S L A D E Y E Y M R K	(SEQ ID NO: 7)
	A G Y F A E X A R	(SEQ ID NO: 11)
	T T E M A S E Q G A	(SEQ ID NO: 13)
	A K E A L A A L K	(SEQ ID NO: 14)
20	F V L Q A K K	(SEQ ID NO: 15)
	E T Q P D P G Q I L K K V P M V I G A Y T	(SEQ ID NO: 165)
	E Y K C L K F K W F K K A T V M	(SEQ ID NO: 17)
	E X K F Y V P	(SEQ ID NO: 19)
25	K L E F L X A K	(SEQ ID NO: 32); and

(b) a basic polypeptide factor for use in treating muscle cells which is also known to stimulate glial cell mitogenesis in the presence of fetal calf plasma, has a molecular weight of from about 55 kD to
30 about 63 kD, and including within its amino acid sequence any one or more of the following peptide sequences:

- 13 -

	V H Q V W A A K	(SEQ ID NO: 33)
	Y I F F M E P E A X S S G	(SEQ ID NO: 34)
	L G A W G P P A F P V X Y	(SEQ ID NO: 35)
	W F V V I E G K	(SEQ ID NO: 36)
5	A S P V S V G S V Q E L Q R	(SEQ ID NO: 37)
	V C L L T V A A L P P T	(SEQ ID NO: 38)
	K V H Q V W A A K	(SEQ ID NO: 48)
	K A S L A D S G E Y M X K	(SEQ ID NO: 49)
	D L L L X V	(SEQ ID NO: 39)

10 Methods for the use of the peptide sequences set out above, derived from the smaller molecular weight polypeptide factor, and from the larger molecular weight polypeptide factor, are also aspects of this invention. Monoclonal antibodies to the above peptides are

15 themselves useful investigative tools and therapeutics.

Thus, the invention further embraces methods of using a polypeptide factor having activities useful for treating muscle cells and including an amino acid sequence encoded by:

20 (a) a DNA sequence shown in any one of Figs. 27A, 27B or 27C, SEQ ID Nos. 129-131, respectively;

 (b) a DNA sequence shown in Fig. 21, SEQ ID No. 85;

 (c) the DNA sequence represented by nucleotides 25 281-557 of the sequence shown in Fig. 27A, SEQ ID No. 129; or

 (d) a DNA sequence hybridizable to any one of the DNA sequences according to (a), (b) or (c).

Following factors as muscle cell mitogens:

30 (a) a basic polypeptide factor which has, if obtained from bovine pituitary material, an observed molecular weight, whether in reducing conditions or not, of from about 30kD to about 36kD on SDS-polyacrylamide

- 14 -

gel electrophoresis which factor has muscle cell mitogenic activity including stimulating the division of myoblasts, and when isolated using reversed-phase HPLC retains at least 50% of said activity after 10 weeks
5 incubation in 0.1% trifluoroacetic acid at 4°C; and

(b) a basic polypeptide factor which has, if obtained from bovine pituitary material, an observed molecular weight, under non-reducing conditions, of from about 55 kD to about 63 Kd on SDS-polyacrylamide gel
10 electrophoresis which factor the human equivalent of which is encoded by DNA clone GGF2HBS5 and which factor has muscle cell mitogenic activity and when isolated using reversed-phase HPLC retains at least 50% of the activity after 4 days incubation in 0.1% trifluoroacetic
15 acid at 4°C.

Thus other important aspects of the invention are the use of:

(a) A series of human and bovine polypeptide factors having cell mitogenic activity including
20 stimulating the division of muscle cells. These peptide sequences are shown in Figs. 30, 31, 32 and 33, SEQ ID Nos. 132-133, respectively.

(b) A series of polypeptide factors having cell mitogenic activity including stimulating the division of
25 muscle cells and purified and characterized according to the procedures outlined by Lupu et al. Science 249: 1552 (1990); Lupu et al. Proc. Natl. Acad. Sci USA 89: 2287 (1992); Holmes et al. Science 256: 1205 (1992); Peles et al. 69: 205 (1992); Yarden and Peles Biochemistry 30:
30 3543 (1991); Dobashi et al. Proc. Natl. Acad. Sci. 88: 8582 (1991); Davis et al. Biochem. Biophys. Res. Commun. 179: 1536 (1991); Beaumont et al., patent application PCT/US91/03443 (1990); Bottenstein, U.S. Patent No.

- 15 -

5,276,145, issued 1/4/94; and Greene et al. patent application PCT/US91/02331 (1990).

(c) A polypeptide factor (GGFBPP5) having glial cell mitogenic activity including stimulating the
5 division of muscle cells. The amino acid sequence is shown in Fig. 31, SEQ ID No. 144.

Methods for stimulating mitogenesis of a myoblast by contacting the myoblast cell with a polypeptide defined above as a muscle cell mitogen *in vivo* or *in*
10 *vitro* are included as features of the invention.

Muscle cell treatments may also be achieved by administering DNA encoding the polypeptide compounds described above in an expressible genetic construction. DNA encoding the polypeptide may be administered to the
15 patient using techniques known in the art for delivering DNA to the cells. For example, retroviral vectors, electroporation or liposomes may be used to deliver DNA.

The invention includes the use of the above named family of proteins as extracted from natural sources
20 (tissues or cell lines) or as prepared by recombinant means.

Other compounds in particular, peptides, which bind specifically to the p185^{erbB2} receptor can also be used according to the invention as muscle cell mitogens.
25 A candidate compound can be routinely screened for p185^{erbB2} binding, and, if it binds, can then be screened for glial cell mitogenic activity using the methods described herein.

The invention includes use of any modifications or
30 equivalents of the above polypeptide factors which do not exhibit a significantly reduced activity. For example, modifications in which amino acid content or sequence is altered without substantially adversely affecting activity are included. The statements of effect and use

- 16 -

contained herein are therefore to be construed accordingly, with such uses and effects employing modified or equivalent factors being part of the invention.

5 The human peptide sequences described above and presented in Figs. 30, 31, 32 and 33, SEQ ID Nos. 132-146, respectively, represent a series of splicing variants which can be isolated as full length complementary DNAs (cDNAs) from natural sources (cDNA
10 libraries prepared from the appropriate tissues) or can be assembled as DNA constructs with individual exons (e.g., derived as separate exons) by someone skilled in the art.

 The invention also includes a method of making a
15 medicament for treating muscle cells, i.e., for inducing muscular mitogenesis, myogenesis, differentiation, or survival, by administering an effective amount of a polypeptide as defined above. Such a medicament is made by administering the polypeptide with a pharmaceutically
20 effective carrier.

 Another aspect of the invention is the use of a pharmaceutical or veterinary formulation comprising any factor as defined above formulated for pharmaceutical or veterinary use, respectively, optionally together with an
25 acceptable diluent, carrier or excipient and/or in unit dosage form. In using the factors of the invention, conventional pharmaceutical or veterinary practice may be employed to provide suitable formulations or compositions.

30 Thus, the formulations to be used as a part of the invention can be applied to parenteral administration, for example, intravenous, subcutaneous, intramuscular, intraorbital, ophthalmic, intraventricular, intracranial, intracapsular, intraspinal, intracisternal,

- 17 -

intraperitoneal, topical, intranasal, aerosol, scarification, and also oral, buccal, rectal or vaginal administration.

5 The formulations of this invention may also be administered by the transplantation into the patient of host cells expressing the DNA encoding polypeptides which are effective for the methods of the invention or by the use of surgical implants which release the formulations of the invention.

10 Parenteral formulations may be in the form of liquid solutions or suspensions; for oral administration, formulations may be in the form of tablets or capsules; and for intranasal formulations, in the form of powders, nasal drops, or aerosols.

15 Methods well known in the art for making formulations are to be found in, for example, "Remington's Pharmaceutical Sciences." Formulations for parenteral administration may, for example, contain as excipients sterile water or saline, polyalkylene glycols
20 such as polyethylene glycol, oils of vegetable origin, or hydrogenated naphthalenes, biocompatible, biodegradable lactide polymer, or polyoxyethylene-polyoxypropylene copolymers may be used to control the release of the present factors. Other potentially useful parenteral
25 delivery systems for the factors include ethylene-vinyl acetate copolymer particles, osmotic pumps, implantable infusion systems, and liposomes. Formulations for inhalation may contain as excipients, for example, lactose, or may be aqueous solutions containing, for
30 example, polyoxyethylene-9-lauryl ether, glycocholate and deoxycholate, or may be oily solutions for administration in the form of nasal drops, or as a gel to be applied intranasally. Formulations for parenteral administration may also include glycocholate for buccal administration,

- 18 -

methoxysalicylate for rectal administration, or citric acid for vaginal administration.

The present factors can be used as the sole active agents, or can be used in combination with other active ingredients, e.g., other growth factors which could facilitate neuronal survival in neurological diseases, or peptidase or protease inhibitors.

The concentration of the present factors in the formulations of the invention will vary depending upon a number of issues, including the dosage to be administered, and the route of administration.

In general terms, the factors of this invention may be provided in an aqueous physiological buffer solution containing about 0.1 to 10% w/v compound for parenteral administration. General dose ranges are from about 1 mg/kg to about 1 g/kg of body weight per day; a preferred dose range is from about 0.01 mg/kg to 100 mg/kg of body weight per day. The preferred dosage to be administered is likely to depend upon the type and extent of progression of the pathophysiological condition being addressed, the overall health of the patient, the make up of the formulation, and the route of administration.

The polypeptide factors utilized in the methods of the invention can also be used as immunogens for making antibodies, such as monoclonal antibodies, following standard techniques. These antibodies can, in turn, be used for therapeutic or diagnostic purposes. Thus, conditions perhaps associated with muscle diseases resulting from abnormal levels of the factor may be tracked by using such antibodies. In vitro techniques can be used, employing assays on isolated samples using standard methods. Imaging methods in which the antibodies are, for example, tagged with radioactive isotopes which can be imaged outside the body using

- 19 -

techniques for the art of tumor imaging may also be employed.

A further general aspect of the invention is the use of a factor of the invention in the manufacture of a medicament, preferably for the treatment of a muscular disease or disorder. The "GGF2" designation is used for all clones which were previously isolated with peptide sequence data derived from GGF-II protein (i.e., GGF2HBS5, GGF2BPP3) and, when present alone (i.e., GGF2 or rhGGF2), to indicate recombinant human protein encoded by plasmids isolated with peptide sequence data derived from the GGF-II protein (i.e., as produced in insect cells from the plasmid HBS5). Recombinant human GGF from the GGFHBS5 clone is called GGF2, rhGGF2 and GGF2HBS5 polypeptide.

Treating as used herein means any administration of the compounds described herein for the purpose of increasing muscle cell mitogenesis, survival, and/or differentiation, and/or decreasing muscle atrophy and degeneration. Most preferably, the treating is for the purpose of reducing or diminishing the symptoms or progression of a disease or disorder of the muscle cells. Treating as used herein also means the administration of the compounds for increasing or altering the muscle cells in healthy individuals. The treating may be brought about by the contacting of the muscle cells which are sensitive or responsive to the compounds described herein with an effective amount of the compound, as described above. Inhibitors of the compounds described herein may also be used to halt or slow diseases of muscle cell proliferation.

- 20 -

Brief Description of the Drawings

The drawings will first be described.

Drawings

Fig. 1 is a graph showing the results of rhGGF2 in
5 a myoblast mitogenesis assay.

Fig. 2 is a graph showing the effect of rhGGF2 on
the number of nuclei in myotubes.

Fig. 3 is a graph of a survival assay showing the
effect of rhGGF2 on survival of differentiated myotubes.

10 Fig. 4 is a graph of survival assays showing the
effect of rhGGF2 on differentiated myotubes relative to
human platelet derived growth factor, human fibroblast
growth factor, human epidermal growth factor, human
leucocyte inhibitory factor, and human insulin-like
15 growth factors I and II.

Fig. 5 is a graph showing the increased survival
on Duchenne muscular dystrophy cells in the presence of
rhGGF2. Fig. 6 is a graph of increasing human growth
hormone (hGH) expression in C2 cells from an hGH reporter
20 gene under control of the AchR delta subunit
transcriptional control elements. This increase is tied
to the addition of GGF2 to the media.

Fig. 7 is a graph of increasing hGH reporter
synthesis and bungarotoxin (BTX) binding to AchRs
25 following the addition of increasing amounts of GGF2 to
C2 cells.

Figs. 8, 9, 10 and 11 are the peptide sequences
derived from GGF-I and GGF-II, SEQ ID Nos. 1-20, 22-29,
32-50 and 165, (see Examples 11-13 hereinafter).

30 Fig. 9, Panel A, is the sequences of GGF-I
peptides used to design degenerate oligonucleotide probes
and degenerate PCR primers are listed (SEQ ID Nos. 1, 17
and 22-29). Some of the sequences in Panel A were also
used to design synthetic peptides. Panel B is a listing

- 21 -

of the sequences of novel peptides that were too short (less than 6 amino acids) for the design of degenerate probes or degenerate PCR primers (SEQ ID Nos. 17 and 32);

Fig. 11, Panel A, is a listing of the sequences of GGF-II peptides used to design degenerate oligonucleotide probes and degenerate PCR primers (SEQ ID Nos. 42-49). Some of the sequences in Panel A were used to design synthetic peptides. Panel B is a listing of the novel peptide that was too short (less than 6 amino acids) for the design of degenerate probes or degenerate PCR primers (SEQ ID No. 50);

Figs. 12, 13A, 13B, 14, 15, 16, 17, 18, and 19 relate to Example 8, below, and depict the mitogenic activity of factors of the invention;

Figs. 20, 21, 22, 23, 24, 25, 26, and 27 relate to Example 10, below and are briefly described below:

Fig. 20 is a listing of the degenerate oligonucleotide probes (SEQ ID Nos. 51-84) designed from the novel peptide sequences in Figure 7, Panel A and Figure 9, Panel A;

Fig. 21 (SEQ ID No. 85) depicts a stretch of the putative bovine GGF-II gene sequence from the recombinant bovine genomic phage GGF2BG1, containing the binding site of degenerate oligonucleotide probes 609 and 650 (see Figure 18, SEQ ID NOS. 66 and 69, respectively). The figure is the coding strand of the DNA sequence and the deduced amino acid sequence in the third reading frame. The sequence of peptide 12 from factor 2 (bold) is part of a 66 amino acid open reading frame (nucleotides 75272);

Fig. 22 is the degenerate PCR primers (Panel A, SEQ'IS Nos. 86-104) and unique PCR primers (Panel B, SEQ ID Nos. 105-115) used in experiments to isolate segments of the bovine GGF-II coding sequences present in RNA from posterior pituitary;

- 22 -

Fig. 23 depicts of the nine distinct contiguous bovine GGF-II cDNA structures and sequences that were obtained in PCR amplification experiments. The top line of the Figure is a schematic of the coding sequences which contribute to the cDNA structures that were characterized;

Fig. 24 is a physical map of bovine recombinant phage of GGF2BG1. The bovine fragment is roughly 20 kb in length and contains two exons (bold) of the bovine GGF-II gene. Restriction sites for the enzymes XbaI, SpeI, NdeI, EcoRI, KpnI, and SstI have been placed on this physical map. Shaded portions correspond to fragments which were subcloned for sequencing;

Fig. 25 is a schematic of the structure of three alternative gene products of the putative bovine GGF-II gene. Exons are listed A through E in the order of their discovery. The alternative splicing patterns 1, 2 and 3 generate three overlapping deduced protein structures (GGF2BPP1, 2, and 3), which are displayed in the various Figs. 27A, 27B, 27C (described below);

Fig. 26 (SEQ ID Nos. 116-128) is a comparison of the GGF-I and GGF-II sequences identified in the deduced protein sequences shown in Figs. 27A, 27B, 27C (described below) with the novel peptide sequences listed in Figs. 9 and 11. The Figure shows that six of the nine novel GGF-II peptide sequences are accounted for in these deduced protein sequences. Two peptide sequences similar to GGF-I sequences are also found;

Fig. 27 (SEQ ID No. 129) is a listing of the coding strand DNA sequence and deduced amino acid sequence of the cDNA obtained from splicing pattern number 1 in Figure 25. This partial cDNA of the putative bovine GGF-II gene encodes a protein of 206 amino acids in length. Peptides in bold were those identified from

- 23 -

the lists presented in Figs. 9 and 11. Potential glycosylation sites are underlined (along with polyadenylation signal AATAAA);

Fig. 27 (SEQ ID No. 130) is a listing of the
5 coding strand DNA sequence and deduced amino acid sequence of the cDNA obtained from splicing pattern number 2 in Fig. 25. This partial cDNA of the putative bovine GGF-II gene encodes a protein of 281 amino acids in length. Peptides in bold are those identified from
10 the lists presented in Figs. 7 and 9. Potential glycosylation sites are underlined (along with polyadenylation signal AATAAA);

Fig. 27 (SEQ ID No. 131) is a listing of the coding strand DNA sequence and deduced amino acid
15 sequence of the cDNA obtained from splicing pattern number 3 in Fig. 25. This partial cDNA of the putative bovine GGF-II gene encodes a protein of 257 amino acids in length. Peptides in bold are those identified from the lists in Figs. 9 and 11. Potential glycosylation
20 sites are underlined (along with polyadenylation signal AATAAA).

Fig. 28, which relates to Example 16 hereinafter, is an autoradiogram of a cross hybridization analysis of putative bovine GGF-II gene sequences to a variety of
25 mammalian DNAs on a southern blot. The filter contains lanes of EcoRI-digested DNA (5 μ g per lane) from the species listed in the Figure. The probe detects a single strong band in each DNA sample, including a four kilobase fragment in the bovine DNA as anticipated by the physical
30 map in Fig. 24. Bands of relatively minor intensity are observed as well, which could represent related DNA sequences. The strong hybridizing band from each of the other mammalian DNA samples presumably represents the GGF-II homologue of those species.

- 24 -

Fig. 29 is a diagram of representative splicing variants. The coding segments are represented by F, E, B, A, G, C, C/D, C/D', D, D', H, K and L. The location of the peptide sequences derived from purified protein are indicated by "o".

Fig. 30 (SEQ ID Nos. 136-143, 156, 157, 169-178) is a listing of the DNA sequences and predicted peptide sequences of the coding segments of GGF. Line 1 is a listing of the predicted amino acid sequences of bovine GGF, line 2 is a listing of the nucleotide sequences of bovine GGF, line 3 is a listing of the nucleotide sequences of human GGF (heregulin) (nucleotide base matches are indicated with a vertical line) and line 4 is a listing of the predicted amino acid sequences of human GGF/hereregulin where it differs from the predicted bovine sequence. Coding segments E, A' and K represent only the bovine sequences. Coding segment D' represents only the human (heregulin) sequence.

Fig. 31 (SEQ ID No. 144) is the predicted GGF2 amino acid sequence and nucleotide sequence of BPP5. The upper line is the nucleotide sequence and the lower line is the predicted amino acid sequence.

Fig. 32 (SEQ ID No. 145) is the predicted amino acid sequence and nucleotide sequence of GGF2BPP2. The upper line is the nucleotide sequence and the lower line is the predicted amino acid sequence.

Fig. 33 (SEQ ID No. 146) is the predicted amino acid sequence and nucleotide sequence of GGF2BPP4. The upper line is the nucleotide sequence and the lower line is the predicted amino acid sequence.

Fig. 34 (SEQ ID Nos. 147-149) depicts the alignment of two GGF peptide sequences (GGF2BPP4 and GGF2BPP5) with the human EGF (hEGF). Asterisks indicate positions of conserved cysteines.

- 25 -

Fig. 35 depicts the level of GGF activity (Schwann cell mitogenic assay) and tyrosine phosphorylation of a ca. 200kD protein (intensity of a 200 kD band on an autoradiogram of a Western blot developed with an antiphosphotyrosine polyclonal antibody) in response to increasing amounts of GGF.

Fig. 36 is a list of splicing variants derived from the sequences shown in Fig. 30.

Fig. 37 is the predicted amino acid sequence, bottom, and nucleic sequence, top, of EGFL1 (SEQ ID No. 150).

Fig. 38 is the predicted amino acid sequence, bottom, and nucleic sequence, top, of EGFL2 (SEQ ID No. 151).

Fig. 39 is the predicted amino acid sequence, bottom, and nucleic sequence, top, of EGFL3 (SEQ ID No. 152).

Fig. 40 is the predicted amino acid sequence, bottom, and nucleic sequence, top, of EGFL4 (SEQ ID No. 153).

Fig. 41 is the predicted amino acid sequence, bottom, and nucleic sequence, top, of EGFL5 (SEQ ID No. 154).

Fig. 42 is the predicted amino acid sequence, bottom, and nucleic sequence, top, of EGFL6 (SEQ ID No. 159).

Fig. 43 is a scale coding segment map of the clone. T3 refers to the bacteriophage promoter used to produce mRNA from the clone. R = flanking EcoRI restriction enzyme sites. 5' UT refers to the 5' untranslated region. E, B, A, C, C/D', and D refer to the coding segments. O = the translation start site. ^ = the 5' limit of the region homologous to the bovine E

- 26 -

segment (see Example 17) and 3' UT refers to the 3' untranslated region.

Fig. 44 is the predicted amino acid sequence (middle) and nucleic sequence (top) of GGF2HBS5 (SEQ ID No. 21). The bottom (intermittent) sequence represents peptide sequences derived from GGF-II preparations (see Figs. 8, 9).

Fig. 45 (A) is a graph showing the purification of rGGF on cation exchange column by fraction; Fig. 45 (B) is a photograph of a Western blot using fractions as depicted in (A) and a GGFII specific antibody.

Fig. 46 is the sequence of the GGFHBS5, GGFHFB1 and GGFBBP5 polypeptides (SEQ ID NOS: 166, 167, and 168).

Fig. 47 is a map of the plasmid pCDHRFpolyA.

15

Detailed Description

The invention pertains to the use of isolated and purified neuregulin factors and DNA sequences encoding these factors, regulatory compounds which increase the extramuscular concentrations of these factors, and compounds which are mimetics of these factors for the induction of muscle cell mitogenesis, differentiation, and survival of the muscle cells *in vivo* and *in vitro*.

It is evident that the gene encoding GGF/p185^{erbB2} binding neuregulin proteins produces a number of variably-sized, differentially-spliced RNA transcripts that give rise to a series of proteins. These proteins are of different lengths and contain some common peptide sequences and some unique peptide sequences. The conclusion that these factors are encoded by a single gene is supported by the differentially-spliced RNA sequences which are recoverable from bovine posterior pituitary and human breast cancer cells (MDA-MB-231)). Further support for this conclusion derives from the size

- 27 -

range of proteins which act as both mitogens for muscle tissue (as disclosed herein) and as ligands for the p185^{erbB2} receptor (see below).

Further evidence to support the fact that the
5 genes encoding GGF/p185^{erbB2} binding proteins are homologous comes from nucleotide sequence comparison. Holmes et al., (Science 256:1205-1210, 1992) demonstrate the purification of a 45-kilodalton human protein
10 (Heregulin- α) which specifically interacts with the receptor protein p185^{erbB2}. Peles et al. (Cell 69:205 (1992)) and Wen et al. (Cell 69:559 (1992)) describe a complementary DNA isolated from rat cells encoding a protein called "neu differentiation factor" (NDF). The translation product of the NDF cDNA has p185^{erbB2} binding
15 activity. Several other groups have reported the purification of proteins of various molecular weights with p185^{erbB2} binding activity. These groups include Lupu et al. ((1992) Proc. Natl. Acad. Sci. USA 89:2287); Yarden and Peles ((1991) Biochemistry 30:3543); Lupu et al. ((1990) Science 249:1552); Dobashi et al. ((1991) Biochem. Biophys. Res. Comm. 179:1536); and Huang et al. ((1992) J. Biol. Chem. 257:11508-11512).

We have found that p185^{erbB2} receptor binding
25 proteins stimulate muscle cell mitogenesis and hence, stimulates myotube formation (myogenesis). This stimulation results in increased formation of myoblasts and increased formation of myotubes (myogenesis). The compounds described herein also stimulate increased muscle growth, differentiation, and survival of muscle
30 cells. These ligands include, but are not limited to the GGF's, the neuregulins, the heregulins, NDF, and ARIA. As a result of this mitogenic activity, these proteins, DNA encoding these proteins, and related compounds may be administered to patients suffering from traumatic damage

- 28 -

or diseases of the muscle tissue. It is understood that all methods provided for the purpose of mitogenesis are useful for the purpose of myogenesis. Inhibitors of these ligands (such as antibodies or peptide fragments) may be administered for the treatment of muscle derived tumors.

These compounds may be obtained using the protocols described herein (Examples 9-17) and in Holmes et al., *Science* 256: 1205 (1992); Peles et al., *Cell* 69:205 (1992); Wen et al., *Cell* 69:559 (1992); Lupu et al., *Proc. Natl. Acad. Sci. USA* 89:2287 (1992); Yarden and Peles, *Biochemistry* 30:3543 (1991); Lupu et al., *Science* 249:1552 (1990); Dobashi et al., *Biochem. Biophys. Res. Comm.* 179:1536 (1991); Huang et al., *J. Biol. Chem.* 257:11508-11512 (1992); Marchionni et al., *Nature* 362:313, (1993); and in the GGF-III patent, all of which are incorporated herein by reference. The sequences are provided and the characteristics described for many of these compounds. For sequences see Figs. 8-11, 20-27C, 29-34, 36-44, and 46. For protein characteristics see Figs. 12-19, 28 35, 45A and 45B.

Compounds may be assayed for their usefulness in vitro using the methods provided in the examples below. In vivo testing may be performed as described in Example 1 and in Sklar et al., *In Vitro Cellular and Developmental Biology* 27A:433-434, 1991.

Other Embodiments

The invention includes methods for the use of any protein which is substantially homologous to the coding segments in Fig. 30 (SEQ ID Nos. 132-143, 156, 1576-147, 160, and 161) as well as other naturally occurring GGF polypeptides for the purpose of inducing muscle mitogenesis. Also included are the use of: allelic

- 29 -

variations; natural mutants; induced mutants; proteins encoded by DNA that hybridizes under high or low stringency conditions to a nucleic acid naturally occurring (for definitions of high and low stringency see
5 Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6, hereby incorporated by reference); and the use of polypeptides or proteins specifically bound by antisera to GGF polypeptides. The term also includes the use of chimeric polypeptides that
10 include the GGF polypeptides comprising sequences from Fig. 28 for the induction of muscle mitogenesis.

As will be seen from Example 8, below, the present factors exhibit mitogenic activity on a range of cell types. The general statements of invention above in
15 relation to formulations and/or medicaments and their manufacture should clearly be construed to include appropriate products and uses.

A series of experiments follow which provide additional basis for the claims described herein. The
20 following examples relating to the present invention should not be construed as specifically limiting the invention, or such variations of the invention, now known or later developed.

The examples illustrate our discovery that
25 recombinant human GGF2 (rhGGF2) confers several effects on primary human muscle culture. rhGGF2 has significant effects in three independent biological activity assays on muscle cultures. The polypeptide increased mitogenesis as measured by proliferation of subconfluent
30 quiescent myoblasts, increased differentiation by confluent myoblasts in the presence of growth factor, and increased survival of differentiated myotubes as measured by loss of dye exclusion and increased acetylcholine receptor synthesis. These activities indicate efficacy

- 30 -

of GGF2 and other neuregulins in inducing muscle repair, regeneration, and prophylactic effects on muscle degeneration.

EXAMPLE 1

5 Mitogenic Activity of rhGGF on Myoblasts

Clone GGF2HBS5 was expressed in recombinant Baculovirus infected insect cells as described in Example 14, *infra*, and the resultant recombinant human GGF2 was added to myoblasts in culture (conditioned
10 medium added at 40 μ l/ml). Myoblasts (057A cells) were grown to preconfluence in a 24 well dish. Medium was removed and replaced with DMEM containing 0.5% fetal calf serum with or without GGF2 conditioned medium at a concentration of 40 μ l/ml. Medium was changed after 2
15 days and cells were fixed and stained after 5 days. Total nuclei were counted as were the number of nuclei in myoblasts (Table 1).

TABLE 1

Treatment	Total Number of Nuclei/mm ²	Nuclei in Myotubes	Fusion Index
20 Control	395 \pm 28.3	204 \pm 9.19	0.515 \pm 0.01
GGF 40 μ l/ml	636 \pm 8.5	381 \pm 82.7	0.591 \pm 0.15

GGF treated myoblasts showed an increased number of total nuclei (636 nuclei) over untreated controls (395 nuclei) indicating mitogenic activity. rhGGF2 treated myotubes
25 had a greater number of nuclei (381 nuclei) than untreated controls (204 nuclei). Thus, rhGGF2 enhances the total number of nuclei through proliferation and increased cell survival. rhGGF2 is also likely to enhance the formation of myotubes.

- 31 -

The mitogenic activity of rhGGF2 may be measured *in vivo* by giving a continuous supply of GGF2 and [³H]thymidine to rat muscle via an osmotic mini pump. The muscle bulk is determined by wet weight after one and
5 two weeks of treatment. DNA replication is measured by counting labeled nuclei in sections after coating for autoradiography (Sklar et al., *In Vitro Cellular and Developmental Biology* 27A:433-434, 1991) in sham and rhGGF2-treated muscle. Denervated muscle is also
10 examined in this rat animal model via these methods and this method allows the assessment of the role of rhGGF2 in muscle atrophy and repair. Mean fiber diameter can also be used for assessing effects of FGF on prevention of atrophy.

15

EXAMPLE 2Effect of rhGGF2 on Muscle Cell Mitogenesis

Quiescent primary clonal human myoblasts were prepared as previously described (Sklar, R., Hudson, A., Brown, R., *In vitro Cellular and Developmental Biology* 1991; 27A:433-434). The quiescent cells were treated
20 with the indicated agents (rhGGF2 conditioned media, PDGF with and without methylprednisolone, and control media) in the presence of 10 μ M BrdU, 0.5% FCS in DMEM. After two days the cells were fixed in 4% paraformaldehyde in
25 PBS for 30 minutes, and washed with 70% ethanol. The cells were then incubated with an anti-BrdU antibody, washed, and antibody binding was visualized with a peroxidase reaction. The number of staining nuclei were then quantified per area. The results show that GGF2
30 induces an increase in the number of labelled nuclei per area over controls (see Table 2).

- 32 -

TABLE 2

Mitogenic Effects of GGF on Human Myoblasts

Treatment	Labelled Nuclei/cm ²	T-Test p value
Control	120 ± 22.4	
Infected Control	103 ± 11.9	
GGF 5 µl/ml	223 ± 33.8	0.019
PDGF 20 ng/ml	418 ± 45.8	0.0005
IGFI 30 ng/ml	280 ± 109.6	0.068
Methylprednisolone 1.0 µM	142 ± 20.7	0.293

10 Platelet derived growth factor (PDGF) was used as a positive control. Methylprednisolone (a corticosteroid) was also used in addition to rhGGF2 and showed no significant increase in labelling of DNA.

rhGGF2 purified to homogeneity (>95% pure) is also
 15 mitogenic for human myoblasts (Fig. 1).

Recombinant human GGF2 also causes mitogenesis of primary human myoblasts (see Table 2 and Fig. 1). The mitogenesis assay is performed as described above. The mitotic index is then calculated by dividing the number
 20 of BrdU positive cells by the total number of cells.

EXAMPLE 3Effect of rhGGF2 on Muscle Cell Differentiation

The effects of purified rhGGF2 (95% pure) on muscle culture differentiation were examined (Fig. 2).
 25 Confluent myoblast cultures were induced to differentiate by lowering the serum content of the culture medium from 20% to 0.5%. The test cultures were treated with the indicated concentration of rhGGF2 for six days, refreshing the culture medium every 2 days. The cultures

- 33 -

were then fixed, stained, and the number of nuclei counted per millimeter. The data in Fig. 2 demonstrate a large increase in the number of nuclei in myotubes when rhGGF2 is present, relative to controls.

5

EXAMPLE 4Effect of rhGGF2 on the Survival of Differentiated Myotubes

The survival of differentiated myotubes was significantly increased by rhGGF2 treatment. Muscle
10 cultures were differentiated in the presence of rhGGF2 and at various times the number of dead myotubes were counted by propidium iodide staining. As can be seen in Fig. 3, the number of dead myotubes is lower in the rhGGF2 treated culture at 4, 5, 6, and 8 days of
15 differentiation. The number of nuclei in myotubes was significantly increased by GGF2 treatment compared to untreated cultures after 8 days of differentiation. Specifically, the control showed 8.6 myonuclei/mm², while rhGGF2 treated cultures showed 57.2 myonuclei/mm²
20 (p=0.035) when counted on the same plates after geimsa staining.

The survival assay was also performed with other growth factors which have known effects on muscle culture. The rhGGF2 effect was unique among the growth
25 factors tested (Fig. 4). In this experiment cultures were treated in parallel with the rhGGF2 treated plates with the indicated concentrations of the various growth factors. Survival of myotubes was measured as above at 8 days of differentiation of 057A myoblast cells.
30 Concentrations of factors were as follows: rhGGF2: 100ng/ml; human platelet derived growth factor: 20ng/ml; human basic fibroblast growth factor: 25ng/ml; human epidermal growth factor: 30ng/ml; human leucocyte

- 34 -

inhibitory factor: 10ng/ml; human insulin like growth factor I: 30ng/ml; human insulin like growth factor II: 25ng/ml.

The observed protection of differentiated myotubes from death indicates that rhGGF2 has promise as a therapy for intervention of muscle degeneration characterized by numerous muscle diseases. Thus, agents which increase the extramuscular concentration of neuregulins may have a prophylactic effect or slow the progress of muscle-wasting disorders and increase rates of muscle differentiation, repair, conditioning, and regeneration.

EXAMPLE 5

rhGGF2 Promotes Survival of Differentiated Myotubes with a Genetic Defect at the Duchenne Muscular Dystrophy Locus

The positive effects of rhGGF2 on myotube survival could reflect potential efficacy in degenerative disorders. These effects on myotube survival were tested on a clonally-derived primary Duchenne myoblast to determine if the response observed in normal muscle culture could also be demonstrated in cultures derived from diseased individuals. The data presented in Fig. 5 was obtained using the same muscle culture conditions (Example 4, above) used for normal individual. rhGGF2 significantly decreased the number of dead myotubes in the differentiated Duchenne muscle culture, compared to controls ($p=0.032$). Concentrations were as follows: GGF2: 100ng/ml; human platelet derived growth factor: 20ng/ml; human insulin like growth factor I: 30ng/ml.

This example demonstrates that rhGGF2 can also promote survival of differentiated Duchenne myotubes and provides strong evidence that rhGGF2 may slow or prevent the course of muscle degeneration and wasting in mammals.

- 35 -

EXAMPLE 6rhGGF2 Effect on the Differentiation Program: Induction of MHC Slow and Dystrophin Proteins

The effects of purified rhGGF2 on muscle culture differentiation was also examined by Western analysis of culture lysates. The levels of muscle specific proteins were determined in triplicate treated and untreated cultures. These cultures were prepared and treated as above except that the plate size was increased to 150 mm and the muscle culture layer was scraped off for Western analysis as described in Sklar, R., and Brown, R. (*J. Neurol. Sci.* 101:73-81, 1991). The results presented in Table A indicate that rhGGF2 treatment increases the levels of several muscle specific proteins, including dystrophin, myosin heavy chain (MHC, adult slow and fast isoforms), but does not increase the levels of HSP72 or MHC neonate isoform to a similar level per amount of protein loaded on the Western. The levels of muscle specific proteins induced by rhGGF2 were similar to the quantitative increases in the number of myonuclei/mm² (Table 3).

- 36 -

TABLE 3

value		Control \pm SD	rhGGF2 Treatment \pm SD	p
5	Total Protein (μ g)	554 \pm 38.4	798 \pm 73.6	0.007
	Myonuclei/mm ²	29.0 \pm 12.2	106 \pm 24.1	0.008
	MHC fast/ μ g protein	1.22 \pm 0.47	4.00 \pm 0.40	0.001
	MHC slow/ μ g protein	0.17 \pm 0.13	1.66 \pm 0.27	0.001
	MHC neonate/ μ g protein	0.30 \pm 0.27	0.55 \pm 0.04	0.199
10	dystrophin/ μ g protein	6.67 \pm 0.37	25.5 \pm 11.0	0.042
	HSP 72/ μ g protein	3.30 \pm 0.42	4.54 \pm 0.08	0.008

The rhGGF2 dependent increase in the adult myosin heavy chain isoforms (slow is found in type I human muscle fibers; fast is found in type 2A and 2B human muscle fibers) may represent a maturation of the myotubes, as the neonatal isoform was not significantly increased by rhGGF2 treatment. During rat muscle development MHC isoforms switch from fetal to neonatal forms followed by a switch to mature adult slow and fast MHC isoforms (Periasamy et al. *J. Biol. Chem.* 259:13573-13578, 1984; Periasamy et al. *J. Biol. Chem.* 260:15856-15862, 1985; Wieczorek et al. *J. Cell Biol.* 101:618-629, 1985). While muscle can autonomously undergo some of these isoform transitions in the absence of neural cells or tissue, mouse muscle explants express the adult fast MHC isoform only when cultured in the presence of mouse spinal cord (Ecob-Prince et al. *J. Cell Biol.* 103:995-1005, 1986). Additional evidence that MHC isoform transitions are influenced by nerve was established by Whalen et al. (*Deve. Biol.* 141:24-40, 1990); after regeneration of notexin treated rat soleus muscles only

- 37 -

the adult fast MHC isoform was produced in the new denervated muscle, but innervated regenerated muscle made both fast and slow adult MHC isoforms. Thus the demonstration in Table 3 that rhGGF2 increases the synthesis of adult MHC isoforms indicates that rhGGF2 may induce a developmental maturation of muscle which may mimic neuronal innervation.

EXAMPLE 7

Neuregulins, including rhGGF2, induce the synthesis of acetylcholine receptors in muscle.

The expression of acetylcholine receptor (AChR) subunit proteins can be induced by exposing muscle cells to neuregulins. More specifically, we have shown that contacting muscle cells with rhGGF2 can induce the synthesis of AChR subunit proteins. This induction following rhGGF2 exposure was observed in two ways: first, we detected increased expression of human growth hormone via the product of a reporter gene construct and second we detected increased binding of alpha-bungarotoxin to cells.

In the following example a mouse myoblast cell line C2 was used. C2 cells were transfected with a transgene that contained the 5' regulatory sequences of the AChR delta subunit gene of mouse linked to a human growth hormone full-length cDNA (Baldwin and Burden, 1988. J. Cell Biol. 107:2271-2279). This reporter construct allows the measurement of the induction of AChR delta gene expression by assaying the quantity of growth hormone secreted into the media. The line can be induced to form myotubes by lowering serum concentration in the media from 20% to 0.5%.

Specifically, mouse C2 myoblasts transfected with an AChR-human growth hormone reporter construct and were

- 38 -

assayed for expression of hGH following treatment with rhGGF2. The results of two separate experiments are summarized in Table 4 and in Figures 6 (hGH expression) and 7 (hGH expression and alpha-bungarotoxin binding).

- 5 Shown are the dose response curves for secreted human growth hormone and for bungarotoxin binding from muscle cultures treated with rhGGF2.

TABLE 4

10 Effects of rhGGF2 on the expression of AChR delta subunit/hGH transgene and the synthesis of AChR

	Exp 1		Exp 2	
	GGF (ul)	hGH (ng/ml)	hGH (ng/ml)	AChR (cpm/mg protein)
15	0	9.3 + 2.1	5.7 + 2.1	822 + 170
	0.1	-	6.8 + 1.5	891 + 134
	0.5	-	12.0 + 0.9	993 + 35
	1.0	-	9.7 + 2.3	818 + 67
20	5.0	17.5 + 2.8	14.7 + 3.5	1300 + 177
	10.0	14.3 + 3.2	14.1 + 3.3	1388 + 137
	15.0	22.0 + 1.4	-	-
25	-	-	-	-

- 30 C2 myotubes were treated with cold α -BTX (20 nM) for 1 hr. at 37°C, washed with culture medium twice and then treated with GGF2. Culture medium was adjusted with bovine serum albumin at the concentration of 1 mg/ml. 24 hours later, culture medium was removed and saved for hGH

- 39 -

assay. Muscle cultures were treated with ^{125}I - α -BTX (20 nM) for 1 hour at 37°C, washed and scraped in PBS containing 1% SDS. Non-specific binding was determined in the presence of cold α -BTX (40 nM). The cell
5 homogenate was counted for radioactivity and assayed for total protein amount.

The presence of rhGGF2 led to a greater than 2-fold increase in hGH gene expression, thereby indicating that rhGGF2 induced the synthesis of the delta subunit of
10 the acetylcholine receptor. Furthermore, increased bungarotoxin binding is consistent with assembly of these subunit proteins into functional acetylcholine receptors. To strengthen the interpretation of these data the analysis was repeated on cultures that had the hGH
15 reporter linked to a metallothionein promoter, which should not be responsive to rhGGF2. The results of that control experiment showed that the hGH response was mediated through transcriptional activation of the AchR delta subunit gene control elements.

20 These results indicate that rhGGF2 could be useful in replenishing AchRs as part of the therapy for the autoimmune disease Myasthenia gravis. This activity may also be beneficial in treatment of peripheral nerve regeneration and neuropathy by stimulating a key step in
25 re-innervation of muscle.

EXAMPLE 8

Additional Mitogenic Activities of Purified GGF-I and GGF-II

The mitogenic activity of a highly purified sample
30 containing both GGFs I and II was studied using a quantitative method, which allows a single microculture to be examined for DNA synthesis, cell morphology, cell number and expression of cell antigens. This technique

- 40 -

has been modified from a method previously reported by Muir et al., Analytical Biochemistry 185, 377-382, 1990. The main modifications are: 1) the use of uncoated microtiter plates, 2) the cell number per well, 3) the use of 5% Foetal Bovine Plasma (FBP) instead of 10% Foetal Calf Serum (FCS), and 4) the time of incubation in presence of mitogens and bromodeoxyuridine (BrdU), added simultaneously to the cultures. In addition the cell monolayer was not washed before fixation to avoid loss of cells, and the incubation time of monoclonal mouse anti-BrdU antibody and peroxidase conjugated goat anti-mouse immunoglobulin (IgG) antibody were doubled to increase the sensitivity of the assay. The assay, optimized for rat sciatic nerve Schwann cells, has also been used for several cell lines, after appropriate modifications to the cell culture conditions.

I. Methods of Mitogenesis Testing

On day 1, purified Schwann cells were plated onto uncoated 96 well plates in 5% FBP/Dulbecco's Modified Eagle Medium (DMEM) (5,000 cells/well). On day 2, GGFs or other test factors were added to the cultures, as well as BrdU at a final concentration of 10 μ M. After 48 hours (day 4) BrdU incorporation was terminated by aspirating the medium and cells were fixed with 200 μ l/well of 70% ethanol for 20 min at room temperature. Next, the cells were washed with water and the DNA denatured by incubation with 100 μ l 2N HCl for 10 min at 37°C. Following aspiration, residual acid was neutralized by filling the wells with 0.1 M borate buffer, pH 9.0, and the cells were washed with phosphate buffered saline (PBS). Cells were then treated with 50 μ l of blocking buffer (PBS containing 0.1% Triton X 100 and 2% normal goat serum) for 15 min at 37°C. After

- 41 -

aspiration, monoclonal mouse anti-BrdU antibody (Dako Corp., Santa Barbara, CA) (50 μ l/well, 1.4 μ g/ml diluted in blocking buffer) was added and incubated for two hours at 37°C. Unbound antibodies were removed by three washes

5 in PBS containing 0.1% Triton X-100 and peroxidase-conjugated goat anti-mouse IgG antibody (Dako Corp., Santa Barbara, CA) (50 μ l/well, 2 μ g/ml diluted in blocking buffer) was added and incubated for one hour at 37°C. After three washes in PBS/Triton and a final rinse

10 in PBS, wells received 100 μ l/well of 50 mM phosphate/citrate buffer, pH 5.0, containing 0.05% of the soluble chromogen o-phenylenediamine (OPD) and 0.02% H₂O₂. The reaction was terminated after 5-20 min at room temperature, by pipetting 80 μ l from each well to a clean

15 plate containing 40 μ l/well of 2N sulfuric acid. The absorbance was recorded at 490nm using a plate reader (Dynatech Labs). The assay plates containing the cell monolayers were washed twice with PBS and immunocytochemically stained for BrdU-DNA by adding 100

20 μ l/well of the substrate diaminobenzidine (DAB) and 0.02% H₂O₂ to generate an insoluble product. After 10-20 min the staining reaction was stopped by washing with water, and BrdU-positive nuclei observed and counted using an inverted microscope. occasionally, negative nuclei were

25 counterstained with 0.001% Toluidine blue and counted as before.

II. Cell lines used for Mitogenesis Assays

Swiss 3T3 Fibroblasts: Cells, from Flow Labs, were maintained in DMEM supplemented with 10% FCS, penicillin and streptomycin, at 37°C in a humidified

30 atmosphere of 10% CO₂ in air. Cells were fed or subcultured every two days. For mitogenic assay, cells were plated at a density of 5,000 cells/well in complete

- 42 -

medium and incubated for a week until cells were confluent and quiescent. The serum containing medium was removed and the cell monolayer washed twice with serum free-medium. 100 μ l of serum free medium containing
5 mitogens and 10 μ M of BrdU were added to each well and incubated for 48 hours. Dose responses to GGFs and serum or PDGF (as a positive control) were performed.

BHK (Baby Hamster Kidney) 21 C13 Fibroblasts:

Cells from European Collection of Animal Cell Cultures
10 (ECACC), were maintained in Glasgow Modified Eagle Medium (GMEM) supplemented with 5% tryptose phosphate broth, 5% FCS, penicillin and streptomycin, at 37°C in a humidified atmosphere of 5% CO₂ in air. Cells were fed or subcultured every two to three days. For mitogenic
15 assay, cells were plated at a density of 2,000 cell/well in complete medium for 24 hours. The serum containing medium was then removed and after washing with serum free medium, replaced with 100 μ l of 0.1% FCS containing GMEM or GMEM alone. GGFs and FCS or bFGF as positive controls
20 were added, coincident with 10 μ M BrdU, and incubated for 48 hours. Cell cultures were then processed as described for Schwann cells.

C6 Rat Glioma Cell Line: Cells, obtained at passage 39, were maintained in DMEM containing 5% FCS, 5%
25 Horse serum (HS), penicillin and streptomycin, at 37°C in a humidified atmosphere of 10% CO₂ in air. Cells were fed or subcultured every three days. For mitogenic assay, cells were plated at a density of 2,000 cells/well in complete medium and incubated for 24 hours. Then
30 medium was replaced with a mixture of 1:1 DMEM and F12 medium containing 0.1% FCS, after washing in serum free medium. Dose responses to GGFs, FCS and α FGF were then

- 43 -

performed and cells were processed through the ELISA as previously described for the other cell types.

PC12 (Rat Adrenal Pheochromocytoma Cells): Cells from ECACC, were maintained in RPMI 1640 supplemented with 10% HS, 5% FCS, penicillin and streptomycin, in collagen coated flasks, at 37°C in a humidified atmosphere of 5% CO₂ in air. Cells were fed every three days by replacing 80% of the medium. For mitogenic assay, cells were plated at a density of 3,000 cells/well in complete medium, on collagen coated plates (50 µl/well collagen, Vitrogen Collagen Corp., diluted 1 : 50, 30 min at 37°C) and incubated for 24 hours. The medium was then placed with fresh RPMI either alone or containing 1 mM insulin or 1% FCS. Dose responses to FCS/HS (1:2) as positive control and to GGFs were performed as before. After 48 hours cells were fixed and the ELISA performed as previously described.

III. Results of Mitogenesis Assays: All the experiments presented in this Example were performed using a highly purified sample from a Sepharose 12 chromatography purification step containing a mixture of GGF-I and GGF-II (GGFs).

First, the results obtained with the BrdU incorporation assay were compared with the classical mitogenic assay for Schwann cells based on [125]I-UdR incorporation into DNA of dividing cells, described by J. P. Brockes (*Methods Enzymol.* 147:217, 1987).

Fig. 12 shows the comparison of data obtained with the two assays, performed in the same cell culture conditions (5,000 cells/well, in 5% FBP/DMEM, incubated in presence of GGFs for 48 hrs). As clearly shown, the results are comparable, but BrdU incorporation assay

- 44 -

appears to be slightly more sensitive, as suggested by the shift of the curve to the left of the graph, i.e. to lower concentrations of GGFS.

As described under the section "Methods of Mitogenesis Testing", after the immunoreactive BrdU-DNA has been quantitated by reading the intensity of the soluble product of the OPD peroxidase reaction, the original assay plates containing cell monolayers can undergo the second reaction resulting in the insoluble DAB product, which stains the BrdU positive nuclei. The microcultures can then be examined under an inverted microscope, and cell morphology and the numbers of BrdU-positive and negative nuclei can be observed.

In Fig. 13A and Fig. 13B the BrdU-DNA immunoreactivity, evaluated by reading absorbance at 490 nm, is compared to the number of BrdU-positive nuclei and to the percentage of BrdU-positive nuclei on the total number of cells per well, counted in the same cultures. Standard deviations were less than 10%. The two evaluation methods show a very good correlation and the discrepancy between the values at the highest dose of GGFS can be explained by the different extent of DNA synthesis in cells detected as BrdU-positive.

The BrdU incorporation assay can therefore provide additional useful information about the biological activity of polypeptides on Schwann cells when compared to the (125) I-UdR incorporation assay. For example, the data reported in Fig. 15 show that GGFS can act on Schwann cells to induce DNA synthesis, but at lower doses to increase the number of negative cells present in the microculture after 48 hours.

The assay has then been used on several cell lines of different origin. In Fig. 15 the mitogenic responses of Schwann cells and Swiss 3T3 fibroblasts to GGFS are

- 45 -

compared; despite the weak response obtained in 3T3 fibroblasts, some clearly BrdU-positive nuclei were detected in these cultures. Control cultures were run in parallel in presence of several doses of FCS or human
5 recombinant PDGF, showing that the cells could respond to appropriate stimuli (not shown).

The ability of fibroblasts to respond to GGFs was further investigated using the BHK 21 C13 cell line. These fibroblasts, derived from kidney, do not exhibit
10 contact inhibition or reach a quiescent state when confluent. Therefore the experimental conditions were designed to have a very low background proliferation without compromising the cell viability. GGFs have a significant mitogenic activity on BHK21 C13 cells as
15 shown by Fig. 16 and Fig. 17. Fig. 16 shows the BrdU incorporation into DNA by BHK 21 C13 cells stimulated by GGFs in the presence of 0.1% FCS. The good mitogenic response to FCS indicates that cell culture conditions were not limiting. In Fig. 17 the mitogenic effect of
20 GGFs is expressed as the number of BrdU-positive and BrdU-negative cells and as the total number of cells counted per well. Data are representative of two experiments run in duplicates; at least three fields per well were counted. As observed for Schwann cells in
25 addition to a proliferative effect at low doses, GGFs also increase the numbers of nonresponding cells surviving. The percentage of BrdU positive cells is proportional to the increasing amounts of GGFs added to the cultures. The total number of cells after 48 hours
30 in presence of higher doses of GGFs is at least doubled, confirming that GGFs induce DNA synthesis and proliferation in BHK21 C13 cells. Under the same conditions, cells maintained for 48 hours in the presence

- 46 -

of 2% FCS showed an increase of about six fold (not shown).

C6 glioma cells have provided a useful model to study glial cell properties. The phenotype expressed seems to be dependent on the cell passage, the cells more closely resembling an astrocyte phenotype at an early stage, and an oligodendrocyte phenotype at later stages (beyond passage 70). C6 cells used in these experiments were from passage 39 to passage 52. C6 cells are a highly proliferating population, therefore the experimental conditions were optimized to have a very low background of BrdU incorporation. The presence of 0.1% serum was necessary to maintain cell viability without significantly affecting the mitogenic responses, as shown by the dose response to FCS (Fig. 18).

In Fig. 19 the mitogenic responses to aFGF (acidic Fibroblast growth factor) and GGFs are expressed as the percentages of maximal BrdU incorporation obtained in the presence of FCS (8%). Values are averages of two experiments, run in duplicates. The effect of GGFs was comparable to that of a pure preparation of aFGF. aFGF has been described as a specific growth factor for C6 cells (Lim R. et al., Cell Regulation 1:741-746, 1990) and for that reason it was used as a positive control. The direct counting of BrdU positive and negative cells was not possible because of the high cell density in the microcultures. In contrast to the cell lines so far reported, PC12 cells did not show any evident responsiveness to GGFS, when treated under culture conditions in which PC12 could respond to sera (mixture of FCS and HS as used routinely for cell maintenance). Nevertheless the number of cells plated per well seems to affect the behavior of PC12 cells, and therefore further experiments are required.

- 47 -

EXAMPLE 9Amino acid sequences of purified GGF-I and GGF-II

Amino acid sequence analysis studies were performed using highly purified bovine pituitary GGF-I and GGF-II. The conventional single letter code was used to describe the sequences. Peptides were obtained by lysyl endopeptidase and protease V8 digests, carried out on reduced and carboxymethylated samples, with the lysyl endopeptidase digest of GGF-II carried out on material eluted from the 55-65 RD region of a 11% SDS-PAGE (MW relative to the above-quoted markers).

A total of 21 peptide sequences (see Fig. 8, SEQ ID Nos. 1-20, 165) were obtained for GGF-I, of which 12 peptides (see Fig. 9, SEQ ID Nos. 1, 22-29, 17, 19, and 32) are not present in current protein databases and therefore represent unique sequences. A total of 12 peptide sequences (see Fig. 10, SEQ ID Nos. 42-50 and 161-163) were obtained for GGF-II, of which 10 peptides (see Fig. 11, SEQ ID Nos. 42-50) are not present in current protein databases and therefore represent unique sequences (an exception is peptide GGF-II 06 which shows identical sequences in many proteins which are probably of no significance given the small number of residues). These novel sequences are extremely likely to correspond to portions of the true amino acid sequences of GGFs I and II.

Particular attention can be drawn to the sequences of GGF-I 07 and GGF-II 12, which are clearly highly related. The similarities indicate that the sequences of these peptides are almost certainly those of the assigned GGF species, and are most unlikely to be derived from contaminant proteins.

In addition, in peptide GGF-II 02, the sequence X S S is consistent with the presence of an N linked

- 48 -

carbohydrate moiety on an asparagine at the position denoted by X.

In general, in Figs. 8 and 10, X represents an unknown residue denoting a sequencing cycle where a single position could not be called with certainty either because there was more than one signal of equal size in the cycle or because no signal was present. As asterisk denotes those peptides where the last amino acid called corresponds to the last amino acid present in that peptide. In the remaining peptides, the signal strength after the last amino acid called was insufficient to continue sequence calling to the end of that peptide. The right hand column indicates the results of a computer database search using the GCG package FASTA and TFASTA programs to analyze the NBRF and EMBL sequence databases. The name of a protein in this column denotes identity of a portion of its sequence with the peptide amino acid sequence called allowing a maximum of two mismatches. A question mark denotes three mismatches allowed. The abbreviations used are as follows:

HMG-1	High Mobility Group protein-1
HMG-2	High Mobility Group protein-2
LH-alpha	Luteinizing hormone alpha subunit
LH-beta	Luteinizing hormone beta subunit

25

EXAMPLE 10

Isolating and Cloning of Nucleotide Sequences encoding proteins containing GGF-I and GGF-II peptides

Isolation and cloning of the GGF-II nucleotide sequences was performed as outlined herein, using peptide sequence information and library screening, and was performed as set out below. It will be appreciated that the peptides of Figs. 10 and 11 can be used as the

- 49 -

starting point for isolation and cloning of GGF-I sequences by following the techniques described herein. Indeed, Fig. 20, SEQ ID Nos. 51-84) shows possible degenerate oligonucleotide probes for this purpose, and
5 Fig. 22, SEQ ID Nos. 86-115, lists possible PCR primers. DNA sequence and polypeptide sequence should be obtainable by this means as with GGF-II, and also DNA constructs and expression vectors incorporating such DNA sequence, host cells genetically altered by incorporating
10 such constructs/vectors, and protein obtainable by cultivating such host cells. The invention envisages such subject matter.

I. Design and Synthesis of oligonucleotide Probes and Primers

15 Degenerate DNA oligomer probes were designed by backtranslating the amino acid sequences (derived from the peptides generated from purified GGF protein) into nucleotide sequences. Oligomers represented either the coding strand or the non-coding strand of the DNA
20 sequence. When serine, arginine or leucine were included in the oligomer design, then two separate syntheses were prepared to avoid ambiguities. For example, serine was encoded by either TCN or AGY as in 537 and 538 or 609 and 610. Similar codon splitting was done for arginine or
25 leucine (e.g. 544, 545). DNA oligomers were synthesized on a Biosearch 8750 4-column DNA synthesizer using β -cyanoethyl chemistry operated at 0.2 micromole scale synthesis. Oligomers were cleaved off the column (500 angstrom CpG resins) and deprotected in concentrated
30 ammonium hydroxide for 6-24 hours at 55-60°C. Deprotected oligomers were dried under vacuum (Speedvac) and purified by electrophoresis in gels of 15% acrylamide (20 mono : 1 bis), 50 mM Tris-borate-EDTA buffer

- 50 -

containing 7M urea. Full length oligomers were detected in the gels by UV shadowing, then the bands were excised and DNA oligomers eluted into 1.5 mls H₂O for 4-16 hours with shaking. The eluate was dried, redissolved in 0.1 ml H₂O and absorbance measurements were taken at 260nm.

Concentrations were determined according to the following formula:

$$(A_{260} \times \text{units/ml}) (60.6/\text{length}) = x \mu\text{M}$$

All oligomers were adjusted to 50 μM concentration by addition of H₂O.

Degenerate probes designed as above are shown in Fig. 20, SEQ ID Nos. 54-88.

PCR primers were prepared by essentially the same procedures that were used for probes with the following modifications. Linkers of thirteen nucleotides containing restriction sites were included at the 5' ends of the degenerate oligomers for use in cloning into vectors. DNA synthesis was performed at 1 micromole scale using 1,000 angstrom CpG resins and inosine was used at positions where all four nucleotides were incorporated normally into degenerate probes. Purifications of PCR primers included an ethanol precipitation following the gel electrophoresis purification.

25 II. Library Construction and Screening

A bovine genomic DNA library was purchased from Stratagene (Catalogue Number: 945701). The library contained 2×10^6 15-20kb Sau3A1 partial bovine DNA fragments cloned into the vector lambda DashII. A bovine total brain cDNA library was purchased from Clontech (Catalogue Number: BL 10139). Complementary DNA libraries were constructed (In Vitrogen; Stratagene) from mRNA prepared from bovine total brain, from bovine

- 51 -

pituitary and from bovine posterior pituitary. In Vitrogen prepared two cDNA libraries: one library was in the vector lambda g10, the other in vector pcDNAI (a plasmid library). The Stratagene libraries were prepared
5 in the vector lambda unizap. Collectively, the cDNA libraries contained 14 million primary recombinant phage.

The bovine genomic library was plated on E. coli K12 host strain LE392 on 23 x 23 cm plates (Nunc) at 150,000 to 200,000 phage plaques per plate. Each plate
10 represented approximately one bovine genome equivalent. Following an overnight incubation at 37°C, the plates were chilled and replicate filters were prepared according to procedures of Maniatis et al. (2:60-81). Four plaque lifts were prepared from each plate onto
15 uncharged nylon membranes (Pall Biodyne A or MSI Nitropure). The DNA was immobilized onto the membranes by cross-linking under UV light for 5 minutes or, by baking at 80°C under vacuum for two hours. DNA probes were labelled using T4 polynucleotide kinase (New England
20 Biolabs) with gamma 32P ATP (New England Nuclear; 6500 Ci/mmol) according to the specifications of the suppliers. Briefly, 50 pmols of degenerate DNA oligomer were incubated in the presence of 600 µCi gamma 32P-ATP and 5 units T4 polynucleotide kinase for 30 minutes at
25 37°C. Reactions were terminated, gel electrophoresis loading buffer was added and then radiolabelled probes were purified by electrophoresis. 32P labelled probes were excised from gel slices and eluted into water. Alternatively, DNA probes were labelled via PCR
30 amplification by incorporation of α-32P-dATP or α-32P dCTP according to the protocol of Schowalter and Sommer, Anal. Biochem 177:90-94 (1989). Probes labelled in PCR reactions were purified by desalting on Sephadex G-150 columns.

- 52 -

Prehybridization and hybridization were performed in GMC buffer (0.52 M NaPi, 7% SDS, 1% BSA, 1.5 mM EDTA, 0.1 M NaCl 10 mg/ml tRNA). Washing was performed in oligowash (160 ml 1 M Na₂HPO₄, 200 ml 20% SDS, 8.0 ml 0.5 M EDTA, 100 ml 5M NaCl, 3632 ml H₂O). Typically, 20 filters (400 sq. centimeters each) representing replicate copies of ten bovine genome equivalents were incubated in 200 ml hybridization solution with 100 pmols of degenerate oligonucleotide probe (128-512 fold degenerate). Hybridization was allowed to occur overnight at 5°C below the minimum melting temperature calculated for the degenerate probe. The calculation of minimum melting temperature assumes 2°C for an AT pair and 4°C for a GC pair.

Filters were washed in repeated changes of oligowash at the hybridization temperatures four to five hours and finally, in 3.2M tetramethylammonium chloride, 1% SDS twice for 30 min at a temperature dependent on the DNA probe length. For 20mers, the final wash temperature was 60°C. Filters were mounted, then exposed to X-ray film (Kodak XAR5) using intensifying screens (Dupont Cronex Lightening Plus). Usually, a three to five day film exposure at minus 80°C was sufficient to detect duplicate signals in these library screens. Following analysis of the results, filters could be stripped and reprobed. Filters were stripped by incubating through two successive cycles of fifteen minutes in a microwave oven at full power in a solution of 1% SDS containing 10mM EDTA pH8. Filters were taken through at least three to four cycles of stripping and reprobing with various probes.

- 53 -

III. Recombinant Phage Isolation, Growth and DNA Preparation

These procedures followed standard protocol as described in Recombinant DNA (Maniatis et al 2:60-2:81).

5 IV. Analysis of Isolated Clones Using DNA Digestion and Southern Blots

Recombinant Phage DNA samples (2 micrograms) were digested according to conditions recommended by the restriction endonuclease supplier (New England Biolabs).
10 Following a four hour incubation at 37°C, the reactions products were precipitated in the presence of 0.1M sodium acetate and three volumes of ethanol. Precipitated DNA was collected by centrifugation, rinsed in 75% ethanol and dried. All resuspended samples were loaded onto
15 agarose gels (typically 1% in TAE buffer; 0.04M Tris acetate, 0.002M EDTA). Gel runs were at 1 volt per centimeter from 4 to 20 hours. Markers included lambda Hind III DNA fragments and/or ϕ X174HaeIII DNA fragments (New England Biolabs). The gels were stained with 0.5
20 micrograms/ml of ethidium bromide and photographed. For southern blotting, DNA was first depurinated in the gel by treatment with 0.125 N HCl, denatured in 0.5 N NaOH and transferred in 20x SSC (3M sodium chloride, 0.03 M sodium citrate) to uncharged nylon membranes. Blotting
25 was done for 6 hours up to 24 hours, then the filters were neutralized in 0.5 Tris HCl pH 7.5, 0.15 M sodium chloride, then rinsed briefly in 50 mM Tris-borate EDTA.

For cross-linking, the filters were wrapped first in transparent plastic wrap, then the DNA side exposed
30 for five minutes to an ultraviolet light. Hybridization and washing was performed as described for library screening (see section 2 of this Example). For hybridization analysis to determine whether similar genes

- 54 -

exist in other species slight modifications were made. The DNA filter was purchased from Clontech (Catalogue Number 7753-1) and contains 5 micrograms of EcoRI digested DNA from various species per lane. The probe
5 was labelled by PCR amplification reactions as described in section 2 above, and hybridizations were done in 80% buffer B (2 g polyvinylpyrrolidone, 2 g Ficoll-400, 2 g bovine serum albumin, 50 ml 1M Tris-HCl (pH 7.5) 58 g NaCl, 1 g sodium pyrophosphate, 10 g sodium dodecyl
10 sulfate, 950ml H₂O) containing 10% dextran sulfate. The probes were denatured by boiling for ten minutes then rapidly cooling in ice water. The probe was added to the hybridization buffer at 10⁶ dpm ³²P per ml and incubated overnight at 60°C. The filters were washed at 60°C first
15 in buffer B followed by 2X SSC, 0.1% SDS then in 1x SSC, 0.1% SDS. For high stringency, experiments, final washes were done in 0.1 x SSC, 1% SDS and the temperature raised to 65°C.

Southern blot data were used to prepare a
20 restriction map of the genomic clone and to indicate which subfragments hybridized to the GGF probes (candidates for subcloning).

V. Subcloning of Segments of DNA Homologous to Hybridization Probes

25 DNA digests (e.g. 5 micrograms) were loaded onto 1% agarose gels then appropriate fragments excised from the gels following staining. The DNA was purified by adsorption onto glass beads followed by elution using the protocol described by the supplier (Bio 101). Recovered
30 DNA fragments (100-200 ng) were ligated into linearized dephosphorylated vectors, e.g. pT3T7 (Ambion), which is a derivative of pUC18, using T4 ligase (New England Biolabs). This vector carries the E. coli β lactamase

- 55 -

gene, hence, transformants can be selected on plates containing ampicillin. The vector also supplies β -galactosidase complementation to the host cell, therefore non-recombinants (blue) can be detected using

5 isopropylthiogalactoside and Bluogal (Bethesda Research Labs). A portion of the ligation reactions was used to transform E. coli K12 XL1 blue competent cells (Stratagene Catalogue Number: 200236) and then the

10 transformants were selected on LB plates containing 50 micrograms per ml ampicillin. White colonies were selected and plasmid mini preps were prepared for DNA digestion and for DNA sequence analysis. Selected clones were retested to determine if their insert DNA hybridized with the GGF probes.

15 VI. DNA Sequencing

Double stranded plasmid DNA templates were prepared from 5 ml cultures according to standard protocols. Sequencing was by the dideoxy chain termination method using Sequenase 2.0 and a

20 dideoxynucleotide sequencing kit (US Biochemical) according to the manufacturers protocol (a modification of Sanger et al. PNAS; USA 74:5463 (1977)). Alternatively, sequencing was done in a DNA thermal

25 sequencer (Perkin Elmer, model 4800) using a cycle sequencing kit (New England Biolabs; Bethesda Research Laboratories) and was performed according to manufacturers instructions using a 5'-end labelled

30 primer. Sequence primers were either those supplied with the sequencing kits or were synthesized according to sequence determined from the clones. Sequencing reactions were loaded on and resolved on 0.4mm thick sequencing gels of 6% polyacrylamide. Gels were dried and exposed to X-Ray film. Typically, 35S was

- 56 -

incorporated when standard sequencing kits were used and a 32P end labelled primer was used for cycle sequencing reactions. Sequences were read into a DNA sequence editor from the bottom of the gel to the top (5' direction to 3') and data were analyzed using programs supplied by Genetics Computer Group (GCG, University of Wisconsin).

VII. RNA Preparation and PCR Amplification

Open reading frames detected in the genomic DNA and which contained sequence encoding GGF peptides were extended via PCR amplification of pituitary RNA. RNA was prepared from frozen bovine tissue (Pelfreeze) according to the guanidine neutral-CsCl procedure (Chirgwin et. al. Biochemistry 18:5294(1979).) Polyadenylated RNA was selected by oligo-dT cellulose column chromatography (Aviv and Leder PNAS (USA) 69:1408 (1972)).

Specific DNA target sequences were amplified beginning with either total RNA or polyadenylated RNA samples that had been converted to cDNA using the Perkin Elmer PCR/RNA Kit Number: N808-0017. First strand reverse transcription reactions used 1 µg template RNA and either primers of oligo dT with restriction enzyme recognition site linkers attached or specific antisense primers determined from cloned sequences with restriction sites attached. To produce the second strand, the primers either were plus strand unique sequences as used in 3' RACE reactions (Frohman et. al., PNAS (USA) 85:8998 (1988)) or were oligo dT primers with restriction sites attached if the second target site had been added by terminal transferase tailing first strand reaction products with dATP (e.g. 5' race reactions, Frohman et. al., ibid). Alternatively, as in anchored PCR reactions

- 57 -

the second strand primers were degenerate, hence, representing particular peptide sequences.

The amplification profiles followed the following general scheme: 1) five minutes soak file at 95°C; 2) thermal cycle file of 1 minute, 95°C; 1 minute ramped down to an annealing temperature of 45°C, 50°C or 55°C; maintain the annealing temperature for one minute; ramp up to 72°C over one minute; extend at 72°C for one minute or for one minute plus a 10 second auto extension; 3) extension cycle at 72°C, five minutes, and; 4) soak file 4°C for infinite time. Thermal cycle files (#2) usually were run for 30 cycles. A sixteen μ l sample of each 100 μ l amplification reaction was analyzed by electrophoresis in 2% Nusieve 1% agarose gels run in TAE buffer at 4 volts per centimeter for three hours. The gels were stained, then blotted to uncharged nylon membranes which were probed with labelled DNA probes that were internal to the primers.

Specific sets of DNA amplification products could be identified in the blotting experiments and their positions used as a guide to purification and reamplification. When appropriate, the remaining portions of selected samples were loaded onto preparative gels, then following electrophoresis four to five slices of 0.5 mm thickness (bracketing the expected position of the specific product) were taken from the gel. The agarose was crushed, then soaked in 0.5 ml of electrophoresis buffer from 2-16 hours at 40°C. The crushed agarose was centrifuged for two minutes and the aqueous phase was transferred to fresh tubes.

Reamplification was done on five microliters (roughly 1% of the product) of the eluted material using the same sets of primers and the reaction profiles as in the original reactions. When the reamplification

- 58 -

reactions were completed, samples were extracted with chloroform and transferred to fresh tubes. Concentrated restriction enzyme buffers and enzymes were added to the reactions in order to cleave at the restriction sites present in the linkers. The digested PCR products were purified by gel electrophoresis, then subcloned into vectors as described in the subcloning section above. DNA sequencing was done described as above.

VIII. DNA Sequence Analysis

DNA sequences were assembled using a fragment assembly program and the amino acid sequences deduced by the GCG programs GelAssemble, Map and Translate. The deduced protein sequences were used as a query sequence to search protein sequence databases using WordSearch. Analysis was done on a VAX Station 3100 workstation operating under VMS 5.1. The database search was done on SwissProt release number 21 using GCG Version 7.0.

IX. Results of Cloning and Sequencing of genes encoding GGF-I and GGF-II

As indicated above, to identify the DNA sequence encoding bovine GGF-II degenerate oligonucleotide probes were designed from GGF-II peptide sequences. GGF-II 12 (SEQ ID No. 44), a peptide generated via lysyl endopeptidase digestion of a purified GGF-II preparation (see Figs. 16 and 12) showed strong amino acid sequence homology with GGF-I 07 (SEQ ID No. 39), a tryptic peptide generated from a purified GGF-I preparation. GGF-II 12 was thus used to create ten degenerate oligonucleotide probes (see oligos 609, 610 and 649 to 656 in Fig. 20, SEQ ID Nos. 66, 67, 68 and 75, respectively). A duplicate set of filters were probed with two sets (set 1=609, 610; set 2=649-656) of probes encoding two

- 59 -

overlapping portions of GGF-II 12. Hybridization signals were observed, but, only one clone hybridized to both probe sets. The clone (designated GGF2BG1) was purified.

Southern blot analysis of DNA from the phage clone
5 GGF2BG1 confirmed that both sets of probes hybridized with that bovine DNA sequence, and showed further that both probes reacted with the same set of DNA fragments within the clone. Based on those experiments a 4 kb Eco RI sub-fragment of the original clone was identified,
10 subcloned and partially sequenced. Fig. 21 shows the nucleotide sequence, SEQ ID No. 89) and the deduced amino acid sequence of the initial DNA sequence readings that included the hybridization sites of probes 609 and 650, and confirmed that a portion of this bovine genomic DNA
15 encoded peptide 12 (KASLADSGEYM).

Further sequence analysis demonstrated that GGF-II 12 resided on a 66 amino acid open reading frame (see below) which has become the starting point for the isolation of overlapping sequences representing a
20 putative bovine GGF-II gene and a cDNA.

Several PCR procedures were used to obtain additional coding sequences for the putative bovine GGF-II gene. Total RNA and oligo dT-selected (poly A containing) RNA samples were prepared from bovine total
25 pituitary, anterior pituitary, posterior pituitary, and hypothalamus. Using primers from the list shown in Fig. 22, SEQ ID Nos. 109-119, one-sided PCR reactions (RACE) were used to amplify cDNA ends in both the 3' and 5' directions, and anchored PCR reactions were performed
30 with degenerate oligonucleotide primers representing additional GGF-II peptides. Fig. 29 summarizes the contiguous DNA structures and sequences obtained in those experiments. From the 3' RACE reactions, three alternatively spliced cDNA sequences were produced,

- 60 -

which have been cloned and sequenced. A 5' RACE reaction led to the discovery of an additional exon containing coding sequence for at least 52 amino acids. Analysis of that deduced amino acid sequence revealed

5 peptides GGF-II-6 and a sequence similar to GGF-I-18 (see below). The anchored PCR reactions led to the identification of (cDNA) coding sequences of peptides GGF-II-1, 2, 3 and 10 contained within an additional cDNA segment of 300 bp. The 5' limit of this segment (i.e.,

10 segment E, see Fig. 30) is defined by the oligonucleotide which encodes peptide GGF-II-1 and which was used in the PCR reaction (additional 5' sequence data exists as described for the human clone in Example 11). Thus this clone contains nucleotide sequences encoding six out of

15 the existing total of nine novel GGF-II peptide sequences.

The cloned gene was characterized first by constructing a physical map of GGF2BG1 that allowed us to position the coding sequences as they were found (see

20 below, Fig. 30). DNA probes from the coding sequences described above have been used to identify further DNA fragments containing the exons on this phage clone and to identify clones that overlap in both directions. The putative bovine GGF-II gene is divided into at least 5

25 coding segments. Coding segments are defined as discrete lengths of DNA sequence which can be translated into polypeptide sequences using the universal genetic code. The coding segments described in Fig. 36 and referred to in the present application are: 1) particular exons

30 present within the GGF gene (e.g. coding segment a), or 2) derived from sets of two or more exons that appear in specific sub-groups of mRNAs, where each set can be translated into the specific polypeptide segments as in the gene products shown. The polypeptide segments

- 61 -

referred to in the claims are the translation products of the analogous DNA coding segments. Only coding segments A and B have been defined as exons and sequenced and mapped thus far. The summary of the contiguous coding sequences identified is shown in Fig. 31. The exons are listed (alphabetically) in the order of their discovery. It is apparent from the intron/exon boundaries that exon B may be included in cDNAs that connect coding segment E and coding segment A. That is, exon B cannot be spliced out without compromising the reading frame. Therefore, we suggest that three alternative splicing patterns can produce putative bovine GGF-II cDNA sequences 1, 2 and 3. The coding sequences of these, designated GGF2BPP1.CDS, GGF2BPP2.CDS and GGF2BPP3.CDS, respectively, are given in Figs. 27A (SEQ ID No. 129), 27B (SEQ ID No. 130), and 27C (SEQ ID No. 131), respectively. The deduced amino acid sequence of the three cDNAs is also given in Figs. 27A, (SEQ ID No. 129), 27B (SEQ ID No. 130), and 27C (SEQ ID No. 131).

The three deduced structures encode proteins of lengths 206, 281 and 257 amino acids. The first 183 residues of the deduced protein sequence are identical in all three gene products. At position 184 the clones differ significantly. A codon for glycine GGT in GGF2BPP1 also serves as a splice donor for GGF2BPP2 and GGF2BPP3, which alternatively add on exons C, C/D, C/D' and D or C, C/D and D, respectively, and shown in Fig. 32, SEQ ID No. 145). GGFIIBPP1 is a truncated gene product which is generated by reading past the coding segment A splice junction into the following intervening sequence (intron). This represents coding segment A' in Fig. 30 (SEQ ID No. 136). The transcript ends adjacent to a canonical AATAAA polyadenylation sequence, and we suggest that this truncated gene product represents a

- 62 -

bona fide mature transcript. The other two longer gene products share the same 3' untranslated sequence and polyadenylation site.

5 All three of these molecules contain six of the
nine novel GGF-II peptide sequences (see Fig. 11) and
another peptide is highly homologous to GGF-I-18 (see
Fig. 26). This finding gives a high probability that
this recombinant molecule encodes at least a portion of
10 bovine GGF-II. Furthermore, the calculated isoelectric
points for the three peptides are consistent with the
physical properties of GGF-I and II. Since the molecular
size of GGF-II is roughly 60 kD, the longest of the three
cDNAs should encode a protein with nearly one-half of the
predicted number of amino acids.

15 A probe encompassing the B and A exons was
labelled via PCR amplification and used to screen a cDNA
library made from RNA isolated from bovine posterior
pituitary. One clone (GGF2BPP5) showed the pattern
indicated in Fig. 29 and contained an additional DNA
20 coding segment (G) between coding segments A and C. The
entire nucleic acid sequence is shown in Fig. 31 (SEQ ID
No. 144). The predicted translation product from the
longest open reading frame is 241 amino acids. A portion
of a second cDNA (GGF2BPP4) was also isolated from the
25 bovine posterior pituitary library using the probe
described above. This clone showed the pattern indicated
in Fig. 29. This clone is incomplete at the 5' end, but
is a splicing variant in the sense that it lacks coding
segments G and D. BPP4 also displays a novel 3' end with
30 regions H, K and L beyond region C/D. The sequence of
BPP4 is shown in Fig. 33 (SEQ ID No. 146).

EXAMPLE 11

GGF Sequences in Various Species

- 63 -

The GGF proteins are the members of a new superfamily of proteins. In high stringency cross hybridization studies (DNA blotting experiments) with other mammalian DNAs we have shown, clearly, that DNA probes from this bovine recombinant molecule can readily detect specific sequences in a variety of samples tested. A highly homologous sequence is also detected in human genomic DNA. The autoradiogram is shown in Fig. 28. The signals in the lanes containing rat and human DNA represent the rat and human equivalents of the GGF gene, the sequences of several cDNA's encoded by this gene have been recently reported by Holmes et al. (Science 256: 1205 (1992)) and Wen et al. (Cell 69: 559 (1992)).

EXAMPLE 12

15 Isolation of a Human Sequence Encoding Human GGF2

Several human clones containing sequences from the bovine GGFII coding segment E were isolated by screening a human cDNA library prepared from brain stem (Stratagene catalog #935206). This strategy was pursued based on the strong link between most of the GGF2 peptides (unique to GGF2) and the predicted peptide sequence from clones containing the bovine E segment. This library was screened as described in Example 8, Section II using the oligonucleotide probes 914-919 listed below.

25	914TCGGGCTCCATGAAGAAGATGTA	(SEQ ID NO: 179)
	915TCCATGAAGAAGATGTACCTGCT	(SEQ ID NO: 180)
	916ATGTACCTGCTGTCCTCCTTGA	(SEQ ID NO: 181)
	917TTGAAGAAGGACTCGCTGCTCA	(SEQ ID NO: 182)
	918AAAGCCGGGGCTTGAAGAA	(SEQ ID NO: 183)
30	919ATGARGTGTTGGCGGCGAAA	(SEQ ID NO: 184)

Clones detected with these probes were further analyzed by hybridization. A probe derived from coding segment A (see Fig. 30), which was produced by labeling a

- 64 -

polymerase chain reaction (PCR) product from segment A, was also used to screen the primary library. Several clones that hybridized with both A and E derived probes were selected and one particular clone, GGF2HBS5, was
5 selected for further analysis. This clone is represented by the pattern of coding segments (EBACC/D'D as shown in Fig. 30). The E segment in this clone is the human equivalent of the truncated bovine version of E shown in Fig. 30. GGF2HBS5 is the most likely candidate to encode
10 GGF-II of all the "putative" GGF-II candidates described. The length of coding sequence segment E is 786 nucleotides plus 264 bases of untranslated sequence. The predicted size of the protein encoded by GGF2HBS5 is approximately 423 amino acids (approximately 45
15 kilodaltons, see Fig. 44, SEQ ID NO: 21), which is similar to the size of the deglycosylated form of GGF-II (see Example 20). Additionally, seven of the GGF-II peptides listed in Fig. 26 have equivalent sequences which fall within the protein sequence predicted from
20 region E. Peptides II-6 and II-12 are exceptions, which fall in coding segment B and coding segment A, respectively. RNA encoding the GGF2HBS5 protein was produced in an in vitro transcription system driven by the bacteriophage T7 promoter resident in the vector
25 (Bluescript SK [Stratagene Inc.] see Fig. 47) containing the GGF2HBS5 insert. This RNA was translated in a cell free (rabbit reticulocyte) translation system and the size of the protein product was 45 Kd. Additionally, the cell-free product has been assayed in a Schwann cell
30 mitogenic assay to confirm biological activity. Schwann cells treated with conditioned medium show both increased proliferation as measured by incorporation of ¹²⁵I-Uridine and phosphorylation on tyrosine of a protein in the 185 kilodalton range.

- 65 -

Thus the size of the product encoded by GGF2HBS5 and the presence of DNA sequences which encode human peptides highly homologous to the bovine peptides shown in Fig. 11 confirm that GGF2HBS5 encodes the human
5 equivalent of bovine GGF2. The fact that conditioned media prepared from cells transformed with this clone elicits Schwann cell mitogenic activity confirms that the GGF2HBS5 gene product (unlike the BPP5 gene product) is secreted. Additionally the GGF2HBS5 gene product seems
10 to mediate the Schwann cell proliferation response via a receptor tyrosine kinase such as p185^{erbB2} or a closely related receptor (see Example 19).

EXAMPLE 13

Expression of Human Recombinant GGF2 in Mammalian and 15 Insect Cells

The GGF2HBS5 cDNA clone encoding human GGF2 (as described in Example 12 and also referred to herein as HBS5) was cloned into vector pCDL-SR α 296 and COS-7 cells were transfected in 100 mm dishes by the DEAE-dextran
20 method. Cell lysates or conditioned media from transiently expressing COS cells were harvested at 3 or 4 days post-transfection. To prepare lysates, cell monolayers were washed with PBS, scraped from the dishes lysed by three freeze/thaw cycles in 150 μ m of 0.25 M
25 Tris-HCl, pH8. Cell debris was pelleted and the supernatant recovered. Conditioned media samples (7 mls.) were collected, then concentrated and buffer exchanged with 10 mM Tris, pH 7.4 using Centiprep-10 and Centricon-10 units as described by the manufactures
30 (Amicon, Beverly, MA). Rat nerve Schwann cells were assayed for incorporation of DNA synthesis precursors, as described. Conditioned media or cell lysate samples were tested in the Schwann cell proliferation assay as described in Marchionni et al., Nature 362:313 (1993).

- 66 -

The cDNA, GGF2HBS5, encoding GGF2 directed the secretion of the protein product to the medium. Minimal activity was detectable inside the cells as determined by assays using cell lysates. GGF2HFB1 and GGFBPP5 cDNA's failed to direct the secretion of the product to the extracellular medium. GGF activity from these clones was detectable only in cell lysates.

Recombinant GGF2 was also expressed in CHO cells. The GGF2HBS5 cDNA encoding GGF2 was cloned into the EcoRI site of vector pcdhfrpolyA and transfected into the DHFR negative CHO cell line (GG44) by the calcium phosphate coprecipitation method. Clones were selected in nucleotide and nucleoside free α medium (Gibco) in 96-well plates. After 3 weeks, conditioned media samples from individual clones were screened for expression of GGF by the Schwann cell proliferation assay as described in Marchionni et al., Nature 362:313 (1993). Stable clones which secreted significant levels of GGF activity into the medium were identified. Schwann cell proliferation activity data from different volume aliquots of CHO cell conditioned medium were used to produce the dose response curve shown in Fig. 46 (Graham and Van Der Eb, Virology 52:456, 1973). This material was analyzed on a Western blot probed with polyclonal antisera raised against a GGF2 specific peptide. A band of approximately 65 Kd (the expected size of GGF2 extracted from pituitary) is specifically labeled (Fig. 48, lane 12).

Recombinant GGF2 was also expressed in insect cells using the Baculovirus expression. Sf9 insect cells were infected with baculovirus containing the GGF2HBS5 cDNA clone at a multiplicity of 3-5 (10^6 cells/ml) and cultured in Sf900-II medium. Schwann cell mitogenic activity was secreted into the extracellular medium.

- 67 -

Different volumes of insect cell conditioned medium were tested in the Schwann cell proliferation assay in the absence of forskolin and the data used to produce a dose response curve.

- 5 This material was also analyzed on a Western blot (Fig. 45B) probed with the GGF II specific antibody described above.

 The methods used in this example were as follows:

- Schwann cell mitogenic activity of recombinant
10 human and bovine glial growth factors was determined as follows: Mitogenic responses of cultured Schwann cells were measured in the presence of 5 μ M forskolin using crude recombinant GGF preparations obtained from transient mammalian expression experiments.
15 Incorporation of [125 I]-Urd was determined following an 18-24 hour exposure to materials obtained from transfected or mock transfected cos cells as described in the Methods. The mean and standard deviation of four sets of data are shown. The mitogenic response to
20 partially purified native bovine pituitary GGF (carboxymethyl cellulose fraction; Goodearl et al., submitted) is shown (GGF) as a standard of one hundred percent activity.

- cDNAs (Fig. 46, SEQ ID NOs. 166-168) were cloned
25 into pcDL-SRa296 (Takebe et al., Mol. Cell Biol. 8:466-472 (1988)), and COS-7 cells were transfected in 100 mm dishes by the DEAE-dextran method (Sambrook et al., In *Molecular Cloning. A Laboratory Manual*, 2nd. ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989)). Cell lysates or conditioned media were harvested
30 at 3 or 4 days post-transfection. To prepare lysates, cell monolayers were washed with PBS, scraped from the dishes, and lysed by three freeze/thaw cycles in 150 μ l of 0.25 M Tris-HCl, pH 8. Cell debris was pelleted and

- 68 -

the supernate recovered. Conditioned media samples (7 mls) were collected, then concentrated and buffer exchanged with 10 mM Tris, pH 7.4 using Centriprep-10 and Centricon-10 units as described by the manufacturers (Amicon, Beverly, MA). Rat sciatic nerve Schwann cells were assayed for incorporation of DNA synthesis precursors, as described (Davis and Stroobant, J. Cell Biol. 110:1353-1360 (1990); Brockes et al., Brain Res. 165:105-118 (1979)).

Western blot of recombinant CHO cell conditioned medium were performed as follows: A recombinant CHO clone was cultured in MCDB302 protein-free for 3 days. 2 ml of conditioned medium was harvested, concentrated, buffered exchanged against 10 mM Tris-HCl, pH 7.4 and lyophilized to dryness. The pellet was resuspended in SDS-PAGE sample buffer, subjected to reducing SDS gel electrophoresis and analyzed by Western blotting with a GGF peptide antibody. A CHO control was done by using conditioned medium from untransfected CHO-DG44 host and the CHO HBS5 levels were assayed using conditioned medium from a recombinant clone.

EXAMPLE 14

Identification of Functional Elements of GGF

The deduced structures of the family of GGF sequences indicate that the longest forms (as represented by GGF2BPP4) encode transmembrane proteins where the extracellular part contains a domain which resembles epidermal growth factor (see Carpenter and Wahl in Peptide Growth Factors and Their Receptors I pp. 69-133, Springer-Verlag, NY 1991). The positions of the cysteine residues in coding segments C and C/D or C/D' peptide sequence are conserved with respect to the analogous residues in the epidermal growth factor (EGF) peptide

- 69 -

- sequence (see Fig. 32, SEQ ID Nos. 147-149). This suggests that the extracellular domain functions as receptor recognition and biological activation sites. Several of the variant forms lack the H, K, and L coding segments and thus may be expressed as secreted, diffusible biologically active proteins. GGF DNA sequences encoding polypeptides which encompass the EGF-like domain (EGFL) can have full biological activity for stimulating glial cell mitogenic activity.
- Membrane bound versions of this protein may induce Schwann cell proliferation if expressed on the surface of neurons during embryogenesis or during nerve regeneration (where the surfaces of neurons are intimately associated with the surfaces of proliferating Schwann cells).
- Secreted (non membrane bound) GGFs may act as classically diffusible factors which can interact with Schwann cells at some distance from their point of secretion. Other forms may be released from intracellular sources via tissue injury and cell disruption. An example of a secreted GGF is the protein encoded by GGF2HBS5; this is the only GGF known which has been found to be directed to the exterior of the cell. Secretion is probably mediated via an N-terminal hydrophobic sequence found only in region E, which is the N-terminal domain contained within recombinant GGF2 encoded by GGF2HBS5.
- Other GGF's appear to be non-secreted. These GGFs may be injury response forms which are released as a consequence of tissue damage.
- Other regions of the predicted protein structure of GGF2 (encoded by GGF2HBS5) and other proteins containing regions B and A exhibit similarities to the human basement membrane heparan sulfate proteoglycan core protein. The peptide ADSGEY, which is located next to the second cysteine of the C2 immunoglobulin fold in

- 70 -

these GGF's, occurs in nine of twenty-two C-2 repeats found in that basal lamina protein. This evidence strongly suggests that these proteins may associate with matrix proteins such as those associated with neurons and glia, and may suggest a method for sequestration of glial growth factors at target sites.

EXAMPLE 15

Purification of GGFs from Recombinant Cells

In order to obtain full length or portions of GGFs to assay for biological activity, the proteins can be overproduced using cloned DNA. Several approaches can be used. A recombinant *E. coli* cell containing the sequences described above can be constructed. Expression systems such as pNH8a or pHH16a (Stratagene, Inc.) can be used for this purpose by following manufacturers procedures. Alternatively, these sequences can be inserted in a mammalian expression vector and an overproducing cell line can be constructed. As an example, for this purpose DNA encoding a GGF, clone GGF2BPP5 has been expressed in COS cells and can be expressed in Chinese hamster ovary cells using the pMSXND expression vector (Lee and Nathans, J. Biol. Chem. 263, 3521-3527, (1981)). This vector containing GGF DNA sequences can be transfected into host cells using established procedures.

Transient expression can be examined or G418-resistant clones can be grown in the presence of methotrexate to select for cells that amplify the dhfr gene (contained on the pMSXND vector) and, in the process, co-amplify the adjacent GGF protein encoding sequence. Because CHO cells can be maintained in a totally protein-free medium (Hamilton and Ham, In Vitro 13, 537-547 (1977)), the desired protein can be purified

- 71 -

from the medium. Western analysis using the antisera produced in Example 17 can be used to detect the presence of the desired protein in the conditioned medium of the overproducing cells.

5 The desired protein (rGGF2) was purified from the medium conditioned by transiently expressing cos cells as follows. rGGF II was harvested from the conditioned medium and partially purified using Cation Exchange Chromatography (POROS-HS). The column was equilibrated
10 with 33.3 mM MES pH 6.0. Conditioned media was loaded at flow rate of 10 ml/min. The peak containing Schwann cell proliferation activity and immunoreactive (using the polyclonal antisera was against a GGF2 peptide described above) was eluted with 50 mM Tris, 1M NaCl pH 8.0.

15 rhGGF2 is also expressed using a stable Chinese Ovary Hamster cell line. rGGF2 from the harvested conditioned media was partially purified using Cation Exchange Chromatograph (POROS-HS). The column was equilibrated with PBS pH 7.4. Conditioned media was
20 loaded at 10 ml/min. The peak containing the Schwann Cell Proliferative activity and immunoreactivity (using GGF2 polyclonal antisera) was eluted with 50 mM Hepes, 500 mM NaCl pH 8.0. An additional peak was observed at 50 mM Hepes, 1M NaCl pH 8.0 with both proliferation as
25 well as immunoreactivity (Fig. 45).

 rhGGF2 can be further purified using Hydrophobic Interaction Chromatography as a high resolution step; Cation exchange/Reserve phase Chromatography (if needed as second high resolution step); A viral inactivation
30 step and a DNA removal step such as Anion exchange chromatography.

 Schwann Cell Proliferation Activity of recombinant GGF2 peak eluted from the Cation Exchange column was determined as follows: Mitogenic responses of the

- 72 -

cultured Schwann cells were measured in the presence of 5 M Forskolin using the peak eluted by 50 mM Tris 1 M NaCl pH 8.0. The peak was added at 20 1, 10 1 (1:10) 10 1 and (1:100) 10 1. Incorporation of ^{125}I -Uridine was
5 determined and expressed as (CPM) following an 18-24 hour exposure.

An immunoblot using polyclonal antibody raised against a peptide of GGF2 was carried out as follows: 10 1 of different fractions were ran on 4-12% gradient gels.
10 The gels were transferred on to Nitrocellulose paper, and the nitrocellulose blots were blocked with 5% BSA and probed with GGF2-specific antibody (1:250 dilution). ^{125}I protein A (1:500 dilution, Specific Activity = 9.0/Ci/g) was used as the secondary antibody. The immunoblots were
15 exposed to Kodax X-Ray films for 6 hours. The peak fractions eluted with 1 M NaCl showed an immunoreactive band at 69K.

GGF2 purification on cation exchange columns was performed as follows: CHO cell conditioned media
20 expressing rGGFII was loaded on the cation exchange column at 10 ml/min. The column was equilibrated with PBS pH 7.4. The elution was achieved with 50 mM Hepes 500 mM NaCl pH 8.0 and 50 mM Hepes 1M NaCl pH 8.0 respectively. All fractions were analyzed using the
25 Schwann cell proliferation assay (CPM) described herein. The protein concentration (mg/ml) was determined by the Bradford assay using BSA as the standard.

A Western blot using 10 1 of each fraction was performed and immunoreactivity and the Schwann cell
30 activity were observed to co-migrate.

The protein may be assayed at various points in the procedure using a Western blot assay. Alternatively, the Schwann cell mitogenic assay described herein may be used to assay the expressed product of the full length

- 73 -

clone or any biologically active portions thereof. The full length clone GGF2BPP5 has been expressed transiently in COS cells. Intracellular extracts of transfected COS cells show biological activity when assayed in the

5 Schwann cell proliferation assay described in Example 8. In addition, the full length close encoding GGF2HBS5 has been expressed transiently in COS cells. In this case both cell extract and conditioned media show biological activity in the Schwann cell proliferation assay

10 described in Example 8. Any member of the family of splicing variant complementary DNA's derived from the GGF gene (including the Heregulins) can be expressed in this manner and assayed in the Schwann cell proliferation assay by one skilled in the art.

15 Alternatively, recombinant material may be isolated from other variants according to Wen et al. (Cell 69:559 (1992)) who expressed the splicing variant Neu differentiation factor (NDF) in COS-7 cells. cDNA clones inserted in the pJT-2 eukaryotic plasmid vector

20 are under the control of the SV40 early promoter, and are 3'-flanked with the SV40 termination and polyadenylation signals. COS-7 cells were transfected with the pJT-2 plasmid DNA by electroporation as follows: 6×10^6 cells (in 0.8 ml of DMEM and 10% FBS) were transferred to a

25 0.4 cm cuvette and mixed with 20 μ g of plasmid DNA in 10 μ l of TE solution (10 mM Tris-HCl (pH 8.0), 1 mM EDTA). Electroporation was performed at room temperature at 1600 V and 25 μ F using a Bio-Rad Gene Pulser apparatus with the pulse controller unit set at 200 ohms. The cells

30 were then diluted into 20 ml of DMEM, 10% FBS and transferred into a T75 flask (Falcon). After 14 hr. of incubation at 37°C, the medium was replaced with DMEM, 1% FBS, and the incubation continued for an additional 48 hr. Conditioned medium containing recombinant protein

- 74 -

which was harvested from the cells demonstrated biological activity in a cell line expressing the receptor for this protein. This cell line (cultured human breast carcinoma cell line AU 565) was treated with recombina-
5 nt material. The treated cells exhibited a morphology change which is characteristic of the activation of the erbB2 receptor. Conditioned medium of this type also can be tested in the Schwann cell proliferation assay.

10

EXAMPLE 16N-terminal sequence analysis

The cDNA encoding hGGF2 was cloned into the amplifiable vector pcdhfrpolyA and transfected into CHO-DG44 cells for stable expression. rhGGF2 is secreted
15 into the conditioned media. The ability of the recombinant GGF2 to be secreted is presumably mediated through the N-terminal hydrophobic stretch (signal sequence). A signal sequence, once having initiated export of a growing protein chain across the rough
20 endoplasmic reticulum, is cleaved from the mature protein at a specific site. N-terminal sequence analysis of the expressed and purified rhGGF2 indicates the site of cleavage as shown below. The sequence of the first 50 amino acid residues at the N-terminus of the protein was

- 75 -

confirmed by N-terminal sequence analysis (Table 5),
below.

TABLE 5

N-terminal sequence analysis of rhGGF2

5	Cycle #	Primary Sequence	pMoles
	1	Gly (G)	210.6
	2	Asn (N)	163
	3	GLU (E)	149
	4	Ala (A)	220
10	5	Ala (A)	180
	6	Pro (P)	173
	7	Ala (A)	177
	8	Gly (G)	154.9
	9	Ala (A)	162.4
15	10	Ser (S)	65.4
	11	Val (V)	132.7
	12	Val (V) *(Cys)	11.7
	13	Tyr (Y)	112.7
	14	Ser (S)	47.6
20	15	Ser (S)	27.1

The N-terminal sequence analysis is performed by Edman Degradation Process

*Cys residues are destroyed by the Edman Degradation Process and cannot be detected

25 The following sequence (SEQ ID NO: 185) represents
the amino acid sequence of hGGF2. The shaded area
indicates the cleaved signal sequence.

MRWRAPRRS GRPGPRAQRF GSAARSSPPL PLLPLLELLG TAALAPGAAA
GNEAAPAGAS VCYSSPPSVG SVQELAQRRA VVIEGKVHPQ RRQOGALDRK
30 AAAAAGEAGA WGGDREPPAA GPRALGPPAE EPLLAANGTV PSWPTAPVPS

- 76 -

AGEPGEEAPY LVKVHQVWAV KAGGLKKDSL LTVRLGTWGH PAFPSCGRLK
 EDSRYIFFME PDANSTSRAP AAFRASFPPL ETGRNLKKEV SRVLCKRCAL
 PPQLKEMKSQ ESAAGSKLVL RCETSSEYSS LRFKWFKNNGN ELNRKNKPQN
 IKIQKPGKS ELRINKASLA DSGEYMCKVI SKLGNDASASA NITIVESNAT
 5 STSTTGTS HL VKCAEKEKTF CVNGGECFMV KDLSNPSRYL CKCPNEFTGD
 RCQNYVMASF YSTSTPFLSL PE (SEQ ID NO: 185)

The shaded area represents experimentally determined 15
 amino acid residues at the N-terminal of the rhGGF2,
 indicating A₅₀-G₅₁ bond to be the cleavage site for the
 10 signal sequence.

EXAMPLE 17

Isolation of a Further Splicing Variant

Methods for updating other neuregulins described
 in U.S. patent application Serial No. 07/965,173, filed
 15 October 23, 1992, incorporated herein by reference,
 produced four closely related sequences (heregulin α , β 1,
 β 2, β 3) which arise as a result of splicing variation.
 Peles et al. (Cell 69:205 (1992)), and Wen et al. (Cell
 69:559 (1992)) have isolated another splicing variant
 20 (from rat) using a similar purification and cloning
 approach to that described in Examples 1-9 and 11
 involving a protein which binds to p185^{erbB2}. The cDNA
 clone was obtained as follows (via the purification and
 sequencing of a p185^{erbB2} binding protein from a
 25 transformed rat fibroblast cell line).
 A p185^{erbB2} binding protein was purified from conditioned
 medium as follows. Pooled conditioned medium from three
 harvests of 500 roller bottles (120 liters total) was
 cleared by filtration through 0.2 μ filters and
 30 concentrated 31-fold with a Pelicon ultrafiltration
 system using membranes with a 20kd molecular size cutoff.
 All the purification steps were performed by using a

- 77 -

Pharmacia fast protein liquid chromatography system. The concentrated material was directly loaded on a column of heparin-Sepharose (150 ml, preequilibrated with phosphate-buffered saline (PBS)). The column was washed with PBS containing 0.2 M NaCl until no absorbance at 280 nm wavelength could be detected. Bound proteins were then eluted with a continuous gradient (250 ml) of NaCl (from 0.2 M to 1.0 M), and 5 ml fractions were collected. Samples (0.01 ml of the collected fractions were used for the quantitative assay of the kinase stimulatory activity. Active fractions from three column runs (total volume = 360 ml) were pooled, concentrated to 25 ml by using a YM10 ultrafiltration membrane (Amicon, Danvers, MA), and ammonium sulfate was added to reach a concentration of 1.7 M. After clearance by centrifugation (10,000 x g, 15 min.), the pooled material was loaded on a phenyl-Superose column (HR10/10, Pharmacia). The column was developed with a 45 ml gradient of $(\text{NH}_4)_2\text{SO}_4$ (from 1.7 M to no salt) in 0.1 M Na_2PO_4 (pH 7.4), and 2 ml fractions were collected and assayed (0.002 ml per sample) for kinase stimulation (as described in Example 19). The major peak of activity was pooled and dialyzed against 50 mM sodium phosphate buffer (pH 7.3). A Mono-S cation-exchange column (HR5/5, Pharmacia) was preequilibrated with 50 mM sodium phosphate. After loading the active material (0.884 mg of protein; 35 ml), the column was washed with the starting buffer and then developed at a rate of 1 ml/min. with a gradient of NaCl. The kinase stimulatory activity was recovered at 0.45-0.55 M salt and was spread over four fractions of 2 ml each. These were pooled and loaded directly on a Cu^{+2} chelating columns (1.6 ml, HR2/5 chelating Superose, Pharmacia). Most of the proteins adsorbed to the resin, but they gradually eluted with a

- 78 -

30 ml linear gradient of ammonium chloride (0-1 M). The activity eluted in a single peak of protein at the range of 0.05 to 0.2 M NH_4Cl . Samples from various steps of purification were analyzed by gel electrophoresis followed by silver staining using a kit from ICN (Costa Mesa, CA), and their protein contents were determined with a Coomassie blue dye binding assay using a kit from Bio-Rad (Richmond, CA).

The p44 protein (10 μg) was reconstituted in 200 μl of 0.1 M ammonium bicarbonate buffer (pH 7.8). Digestion was conducted with L-1-tosyl-amide 2-phenylethyl chloromethyl ketone-treated trypsin (Serva) at 37°C for 18 hr. at an enzyme-to-substrate ratio of 1:10. The resulting peptide mixture was separated by reverse-phase HPLC and monitored at 215 nm using a Vydac C4 micro column (2.1 mm i.d. x 15 cm, 300 Å) and an HP 1090 liquid chromatographic system equipped with a diode-array detector and a workstation. The column was equilibrated with 0.1% trifluoroacetic acid (mobile phase A), and elution was effected with a linear gradient from 0%-55% mobile phase B (90% acetonitrile in 0.1% trifluoroacetic acid) over 70 min. The flow rate was 0.2 ml/min. and the column temperature was controlled at 25°C. One-third aliquots of the peptide peaks collected manually from the HPLC system were characterized by N-terminal sequence analysis by Edman degradation. The fraction eluted after 27.7 min. (T27.7) contained mixed amino acid sequences and was further rechromatographed after reduction as follows: A 70% aliquot of the peptide fraction was dried in vacuo and reconstituted in 100 μl of 0.2 M ammonium bicarbonate buffer (pH 7.8). DTT (final concentration 2 mM) was added to the solution, which was then incubated at 37°C for 30 min. The reduced peptide mixture was then separated by reverse-phase HPLC

- 79 -

using a Vydac column (2.1 mm i.d. x 15 cm). Elution conditions and flow rate were identical to those described above. Amino acid sequence analysis of the peptide was performed with a Model 477 protein sequencer (Applied Biosystems, Inc., Foster City, CA) equipped with an on-line phenylthiohydantoin (PTH) amino acid analyzer and a Model 900 data analysis system (Hunkapiller et al. (1986) In Methods of Protein Microcharacterization, J.E. Shively, ed. (Clifton, New Jersey: Humana Press p. 223-247). The protein was loaded onto a trifluoroacetic acid-treated glass fiber disc precycled with polybrene and NaCl. The PTH-amino acid analysis was performed with a micro liquid chromatography system (Model 120) using dual syringe pumps and reverse-phase (C-18) narrow bore columns (Applied Biosystems, 2.1 mm x 250 mm). RNA was isolated from Rat1-EJ cells by standard procedures (Maniatis et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York (1982) and poly (A)⁺ was selected using an mRNA Separator kit (Clontech Lab, Inc., Palo Alto, CA). cDNA was synthesized with the Superscript kit (from BRL Life Technologies, Inc., Bethesda, MD). Column-fractionated double-strand cDNA was ligated into an SalI- and NotI-digested pJT-2 plasmid vector, a derivative of the pCD-X vector (Okayama and Berg, Mol. Cell Biol. 3: 280 (1983)) and transformed into DH10B E. coli cells by electroporation (Dower et al., Nucl. Acids Res. 16: 6127 (1988)). Approximately 5×10^5 primary transformants were screened with two oligonucleotide probes that were derived from the protein sequences of the N-terminus of NDF (residues 5-24) and the T40.4 tryptic peptide (residues 7-12). Their respective sequences were as follows (N indicates all 4 nt):

- 80 -

(1) 5'-ATA GGG AAG GGC GGG GGA AGG GTC NCC CTC NGC

A T

AGG GCC GGG CTT GCC TCT GGA GCC TCT-3'

(2) 5'-TTT ACA CAT ATA TTC NCC-3'

5 C G G C

(1: SEQ ID No. 163; 2: SEQ ID No. 164)

The synthetic oligonucleotides were end-labeled with [γ - 32 P]ATP with T4 polynucleotide kinase and used to screen replicate sets of nitrocellulose filters. The hybridization solution contained 6 x SSC, 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 2 x Denhardt's solution, 50 μ g/ml salmon sperm DNA, and 20% formamide (for probe 1) or no formamide (for probe 2). The filters were washed at either 50°C with 0.5 x SSC, 0.2% SDS, 2 mM EDTA (for probe 1) or at 37°C with 2 x SSC, 0.2% SDS, 2 mM EDTA (for probe 2). Autoradiography of the filters gave ten clones that hybridized with both probes. These clones were purified by replating and probe hybridization as described above.

The cDNA clones were sequenced using an Applied Biosystems 373A automated DNA sequencer and Applied Biosystems Taq DyeDeoxy™ Terminator cycle sequencing kits following the manufacture's instructions. In some instances, sequences were obtained using [35 S]dATP (Amersham) and Sequenase™ kits from U.S. Biochemicals following the manufacturer's instructions. Both strands of the cDNA clone 44 were sequenced by using synthetic oligonucleotides as primers. The sequence of the most 5' 350 nt was determined in seven independent cDNA clones.

The resultant clone demonstrated the pattern shown in Fig. 27 (NDF).

- 81 -

EXAMPLE 19Purification and Assay of Other Proteins which bind
p185^{erbB2} ReceptorI. Purification of gp30 and p70

- 5 Lupu et al. (Science 249, 1552 (1990)) and Lippman and Lupu (patent application number PCT/US91/03443 (1990)), hereby incorporated by reference, have purified a protein from conditioned media of a human breast cancer cell line MDA-MB-231.
- 10 Lupu et al. (Proc. Natl. Acad. Sci. 89, 2287 (1992)) purified another protein which binds to the p185^{erbB2} receptor. This particular protein, p75, was purified from conditioned medium used for the growth of SKBr-3 (a human breast cancer cell line) propagated in
- 15 improved Eagle's medium (IMEM: GIBCO) supplemented with 10% fetal bovine serum (GIBCO).

II. Other p185^{erbB2} ligands

- Peles et al. (Cell 69, 205 (1992)) have also purified a 185^{erbB2} stimulating ligand from rat cells.
- 20 Holmes et al. (Science 256, 1205 (1992)) have purified Heregulin α from human cells which binds and stimulates 185^{erbB2} (see Example 5). Tarakovsky et al. Oncogene 6:218 (1991) have demonstrated binding of a 25 kD polypeptide isolated from activated macrophages to the
- 25 Neu receptor, a p185^{erbB2} homology, herein incorporated by reference.

III. NDF Isolation

- Yarden and Peles (Biochemistry 30, 3543 (1991)) have identified a 35 kilodalton glycoprotein which will
- 30 stimulate the 185^{erbB2} receptor.

- 82 -

In other publications, Davis et al. (Biochem. Biophys. Res. Commun. 179, 1536 (1991), Proc. Natl. Acad. Sci. 88, 8582 (1991) and Greene et al., PCT patent application PCT/US91/02331 (1990)) describe the
5 purification of a protein from conditioned medium of a human T-cell (ATL-2) cell line.

Huang et al. (1992, J. Biol. Chem. 257:11508-11512), hereby incorporated by reference, have isolated
10 an additional neu/erb B2 ligand growth factor from bovine kidney. The 25 kD polypeptide factor was isolated by a procedure of column fractionation, followed by sequential column chromatography on DEAE/cellulose (DE52), Sulfadex (sulfated Sephadex G-50), heparin-Sepharose 4B, and Superdex 75 (fast protein liquid chromatography). The
15 factor, NEL-GF, stimulates tyrosine-specific autophosphorylation of the neu/erb B2 gene product.

IV. Purification of acetylcholine receptor inducing activity (ARIA)

ARIA, a 42 kD protein which stimulates
20 acetylcholine receptor synthesis, has been isolated in the laboratory of Gerald Fischbach (Falls et al., (1993) Cell 72:801-815). ARIA induces tyrosine phosphorylation of a 185 Kda muscle transmembrane protein which resembles p185^{erbB2}, and stimulates acetylcholine receptor synthesis
25 in cultured embryonic myotubes. ARIA is most likely a member of the GGF/erbB2 ligand group of proteins, and this is potentially useful in the glial cell mitogenesis stimulation and other applications of, e.g., GGF2 described herein.

- 83 -

EXAMPLE 19Protein tyrosine phosphorylation mediated by GGF

Rat Schwann cells, following treatment with sufficient levels of Glial Growth Factor to induce proliferation, show stimulation of protein tyrosine phosphorylation. Varying amounts of partially purified GGF were applied to a primary culture of rat Schwann cells according to the procedure outlined in Example 9. Schwann cells were grown in DMEM/10% fetal calf serum/5 μ M forskolin/0.5 μ g per mL GGF-CM (0.5mL per well) in poly D-lysine coated 24 well plates. When confluent, the cells were fed with DMEM/10% fetal calf serum at 0.5mL per well and left in the incubator overnight to quiesce. The following day, the cells were fed with 0.2mL of DMEM/10% fetal calf serum and left in the incubator for 1 hour. Test samples were then added directly to the medium at different concentrations and for different lengths of time as required. The cells were then lysed in boiling lysis buffer (sodium phosphate, 5mM, pH 6.8; SDS, 2%, β -mercapteothanol, 5%; dithiothreitol, 0.1M; glycerol, 10%; Bromophenol Blue, 0.4%; sodium vanadate, 10mM), incubated in a boiling water bath for 10 minutes and then either analyzed directly or frozen at -70°C. Samples were analyzed by running on 7.5% SDS-PAGE gels and then electroblotting onto nitrocellulose using standard procedures as described by Towbin et al. (1979) Proc. Natl. Acad. Sci. USA 76:4350-4354. The blotted nitrocellulose was probed with antiphosphotyrosine antibodies using standard methods as described in Kamps and Selton (1988) Oncogene 2:305-315. The probed blots were exposed to autoradiography film overnight and developed using a standard laboratory processor. Densitometric measurements were carried out using an Ultrascan XL enhanced laser densitometer (LKB).

- 84 -

Molecular weight assignments were made relative to prestained high molecular weight standards (Sigma). The dose responses of protein phosphorylation and Schwann cell proliferation are very similar (Fig. 33). The

5 molecular weight of the phosphorylated band is very close to the molecular weight of p185^{erbB2}. Similar results were obtained when Schwann cells were treated with conditioned media prepared from COS cells translates with the GGF2HBS5 clone. These results correlate well with

10 the expected interaction of the GGFs with and activation of 185^{erbB2}.

This experiment has been repeated with recombinant GGF2. Conditioned medium derived from a CHO cell line stably transformed with the GGF2 clone (GGF2HBS5)

15 stimulates protein tyrosine phosphorylation using the assay described above. Mock transfected CHO cells fail to stimulate this activity.

EXAMPLE 20

N-glycosylation of GGF

20 The protein sequence predicted from the cDNA sequence of GGF-II candidate clones GGF2BPP1,2 and 3 contains a number of consensus N-glycosylation motifs. A gap in the GGFII02 peptide sequence coincides with the asparagine residue in one of these motifs, indicating

25 that carbohydrate is probably bound at this site.

N-glycosylation of the GGFs was studied by observing mobility changes on SDS-PAGE after incubation with N-glycanase, an enzyme that cleaves the covalent linkages between carbohydrate and asparagine residues in

30 proteins.

N-Glycanase treatment of GGF-II yielded a major band of MW 40-42 kDa and a minor band at 45-48 kDa.

- 85 -

Activity single active deglycosylated species at ca 45-50 kDa.

Activity elution experiments with GGF-I also demonstrate an increase in electrophoretic mobility when
5 treated with N-Glycanase, giving an active species of MW 26-28 kDa. Silver staining confirmed that there is a mobility shift, although no N-deglycosylated band could be assigned because of background staining in the sample used.

- 86 -

(1) GENERAL INFORMATION:

- (i) APPLICANTS: Robert Sklar, Mark Marchionni,
David I. Gwynne
- (ii) TITLE OF INVENTION: METHODS FOR ALTERING MUSCLE
CONDITION
- (iii) NUMBER OF SEQUENCES: 185
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Fish & Richardson
(B) STREET: 225 Franklin Street
(C) CITY: Boston
(D) STATE: Massachusetts
(F) ZIP: 02110-2804
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette, 5.25 inch, 360 kb
storage
(B) COMPUTER: IBM
(C) OPERATING SYSTEM: PC-DOS
(D) SOFTWARE: Wordperfect
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/209,204
(B) FILING DATE: 08-MAR-94
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/059,022
(B) FILING DATE: 06-May-93
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Clark, Paul T.
(B) REGISTRATION NUMBER: 30,162
(C) REFERENCE/DOCKET NUMBER: 04585/028WO1
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617) 542-5070
(B) TELEFAX: 200154

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

- 87 -

Phe Lys Gly Asp Ala His Thr Glu
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine; Xaa in position 12 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Xaa Ala Ser Leu Ala Asp Glu Tyr Glu Tyr Met Xaa Lys
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine; Xaa in position 10 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Xaa Thr Glu Thr Ser Ser Ser Gly Leu Xaa Leu Lys
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

- 88 -

Xaa Lys Leu Gly Glu Met Trp Ala Glu
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Xaa Leu Gly Glu Lys Arg Ala
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Xaa Ile Lys Ser Glu His Ala Gly Leu Ser Ile Gly Asp Thr Ala Lys
1 5 10 15

- 89 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Xaa Ala Ser Leu Ala Asp Glu Tyr Glu Tyr Met Arg Lys
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Xaa Ile Lys Gly Glu His Pro Gly Leu Ser Ile Gly Asp Val Ala Lys
1 5 10 15

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine and Xaa in position 12 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Xaa Met Ser Glu Tyr Ala Phe Phe Val Gln Thr Xaa Arg

- 90 -

1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Xaa Ser Glu His Pro Gly Leu Ser Ile Gly Asp Thr Ala Lys
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine; Xaa in position 8 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Xaa Ala Gly Tyr Phe Ala Glu Xaa Ala Arg
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine; Xaa in position 7 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

- 91 -

Xaa Lys Leu Glu Phe Leu Xaa Ala Lys
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Xaa Thr Thr Glu Met Ala Ser Glu Gln Gly Ala
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Xaa Ala Lys Glu Ala Leu Ala Ala Leu Lys
1 5 10

- 92 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Xaa Phe Val Leu Gln Ala Lys Lys
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Xaa Leu Gly Glu Met Trp
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Glu Tyr Lys Cys Leu Lys Phe Lys Trp Phe Lys Lys Ala Thr Val Met
1 5 10 15

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10

- 93 -

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 8 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Glu Ala Lys Tyr Phe Ser Lys Xaa Asp Ala
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 2 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Glu Xaa Lys Phe Tyr Val Pro
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Glu Leu Ser Phe Ala Ser Val Arg Leu Pro Gly Cys Pro Pro Gly Val
1 5 10 15

Asp Pro Met Val Ser Phe Pro Val Ala Leu
20 25

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2003
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 94 -

(ix) FEATURE:

(D) OTHER INFORMATION: N in positions 31 and 32 could be either A or G.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

```

GGAATTCCTT TTTTTTTTTT TTTTTTCTT NNTTTTTTTT TGCCCTTATA CCTCTTCGCC 60
TTTCTGTGGT TCCATCCACT TCTTCCCCCT CCTCCTCCCA TAAACAACTC TCCTACCCCT 120
GCACCCCCAA TAAATAAATA AAAGGAGGAG GGCAAGGGGG GAGGAGGAGG AGTGGTGCTG 180
CGAGGGGAAG GAAAAGGGAG GCAGCGCGAG AAGAGCCGGG CAGAGTCCGA ACCGACAGCC 240
AGAAGCCCGC ACGCACCTCG CACC ATG AGA TGG CGA CGC GCC CCG CGC CGC 291
                Met Arg Trp Arg Arg Ala Pro Arg Arg
                1                    5

TCC GGG CGT CCC GGC CCC CGG GCC CAG CGC CCC GGC TCC GCC GCC CGC 339
Ser Gly Arg Pro Gly Pro Arg Ala Gln Arg Pro Gly Ser Ala Ala Arg
10                15                20                25

TCG TCG CCG CCG CTG CCG CTG CTG CCA CTA CTG CTG CTG CTG GGG ACC 387
Ser Ser Pro Pro Leu Pro Leu Leu Pro Leu Leu Leu Leu Leu Gly Thr
30                35                40

GCG GCC CTG GCG CCG GGG GCG GCG GCC GGC AAC GAG GCG GCT CCC GCG 435
Ala Ala Leu Ala Pro Gly Ala Ala Ala Gly Asn Glu Ala Ala Pro Ala
45                50                55

GGG GCC TCG GTG TGC TAC TCG TCC CCG CCC AGC GTG GGA TCG GTG CAG 483
Gly Ala Ser Val Cys Tyr Ser Ser Pro Pro Ser Val Gly Ser Val Gln
60                65                70

GAG CTA GCT CAG CGC GCC GCG GTG GTG ATC GAG GGA AAG GTG CAC CCG 531
Glu Leu Ala Gln Arg Ala Ala Val Val Ile Glu Gly Lys Val His Pro
75                80                85

CAG CCG CCG CAG CAG GGG GCA CTC GAC AGG AAG GCG GCG GCG GCG GCG 579
Gln Arg Arg Gln Gln Gly Ala Leu Asp Arg Lys Ala Ala Ala Ala Ala
90                95                100                105

GGC GAG GCA GGG GCG TGG GGC GGC GAT CGC GAG CCG CCA GCC GCG GGC 627
Gly Glu Ala Gly Ala Trp Gly Gly Asp Arg Glu Pro Pro Ala Ala Gly
110                115                120

CCA CCG GCG CTG GGG CCG CCC GCC GAG GAG CCG CTG CTC GCC GCC AAC 675
Pro Arg Ala Leu Gly Pro Pro Ala Glu Glu Pro Leu Leu Ala Ala Asn
125                130                135

GGG ACC GTG CCC TCT TGG CCC ACC GCC CCG GTG CCC AGC GCC GGC GAG 723
Gly Thr Val Pro Ser Trp Pro Thr Ala Pro Val Pro Ser Ala Gly Glu
140                145                150

CCC GGG GAG GAG GCG CCC TAT CTG GTG AAG GTG CAC CAG GTG TGG GCG 771
Pro Gly Glu Glu Ala Pro Tyr Leu Val Lys Val His Gln Val Trp Ala
155                160                165

GTG AAA GCC GGG GGC TTG AAG AAG GAC TCG CTG CTC ACC GTG CGC CTG 819
Val Lys Ala Gly Gly Leu Lys Lys Asp Ser Leu Leu Thr Val Arg Leu

```

- 95 -

170	175	180	185	
GGG ACC TGG GGC CAC CCC GCC TTC CCC TCC TGC GGG AGG CTC AAG GAG				867
Gly Thr Trp Gly His Pro Ala Phe Pro Ser Cys Gly Arg Leu Lys Glu	190	195	200	
GAC AGC AGG TAC ATC TTC TTC ATG GAG CCC GAC GCC AAC AGC ACC AGC				915
Asp Ser Arg Tyr Ile Phe Phe Met Glu Pro Asp Ala Asn Ser Thr Ser	205	210	215	
CGC GCG CCG GCC GCC TTC CGA GCC TCT TTC CCC CCT CTG GAG ACG GGC				963
Arg Ala Pro Ala Ala Phe Arg Ala Ser Phe Pro Pro Leu Glu Thr Gly	220	225	230	
CGG AAC CTC AAG AAG GAG GTC AGC CGG GTG CTG TGC AAG CGG TGC GCC				1011
Arg Asn Leu Lys Lys Glu Val Ser Arg Val Leu Cys Lys Arg Cys Ala	235	240	245	
TTG CCT CCC CAA TTG AAA GAG ATG AAA AGC CAG GAA TCG GCT GCA GGT				1059
Leu Pro Pro Gln Leu Lys Glu Met Lys Ser Gln Glu Ser Ala Ala Gly	250	255	260	265
TCC AAA CTA GTC CTT CGG TGT GAA ACC AGT TCT GAA TAC TCC TCT CTC				1107
Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser Ser Leu	270	175	180	
AGA TTC AAG TGG TTC AAG AAT GGG AAT GAA TTG AAT CGA AAA AAC AAA				1155
Arg Phe Lys Trp Phe Lys Asn Gly Asn Glu Leu Asn Arg Lys Asn Lys	185	190	195	
CCA CAA AAT ATC AAG ATA CAA AAA AAG CCA GGG AAG TCA GAA CTT CGC				1203
Pro Gln Asn Ile Lys Ile Gln Lys Lys Pro Gly Lys Ser Glu Leu Arg	200	205	210	
ATT AAC AAA GCA TCA CTG GCT GAT TCT GGA GAG TAT ATG TGC AAA GTG				1251
Ile Asn Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Cys Lys Val	215	220	225	
ATC AGC AAA TTA GGA AAT GAC AGT GCC TCT GCC AAT ATC ACC ATC GTG				1299
Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr Ile Val	230	235	240	245
GAA TCA AAC GCT ACA TCT ACA TCC ACC ACT GGG ACA AGC CAT CTT GTA				1347
Glu Ser Asn Ala Thr Ser Thr Ser Thr Thr Gly Thr Ser His Leu Val	250	255	260	
AAA TGT GCG GAG AAG GAG AAA ACT TTC TGT GTG AAT GGA GGG GAG TGC				1395
Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys	265	270	275	
TTC ATG GTG AAA GAC CTT TCA AAC CCC TCG AGA TAC TTG TGC AAG TGC				1443
Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu Cys Lys Cys	280	285	290	
CCA AAT GAG TTT ACT GGT GAT CGC TGC CAA AAC TAC GTA ATG GCC AGC				1491
Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn Tyr Val Met Ala Ser	295	300	305	
TTC TAC AGT ACG TCC ACT CCC TTT CTG TCT CTG CCT GAA				1530
Phe Tyr Ser Thr Ser Thr Pro Phe Leu Ser Leu Pro Glu	400	405	410	

- 96 -

TAGGAGCATG CTCAGTTGGT GCTGCTTCT TGTGCTGCA TCTCCCTCA GATTCCACCT 1590
 AGAGCTAGAT GTGTCTTACC AGATCTAATA TTGACTGCCT CTGCCTGTCG CATGAGAACA 1650
 TTAACAAAAG CAATTGTATT ACTTCCTCTG TTCGCGACTA GTTGGCTCTG AGATACTAAT 1710
 AGGTGTGTGA GGCTCCGGAT GTTCTTGAA TTGATATTGA ATGATGTGAT ACAAATTGAT 1770
 AGTCAATATC AAGCAGTGAA ATATGATAAT AAAGGCATT CAAAGTCTCA CTTTTATTGA 1830
 TAAAATAAAA ATCATTCTAC TGAACAGTCC ATCTTCTTA TACAATGACC ACATCCTGAA 1890
 AAGGGTGTG CTAAGCTGTA ACCGATATGC ACTTGAAATG ATGGTAAGTT AATTTTGATT 1950
 CAGAAATGTG TATTTGTCAC AAATAAACAT AATAAAAGGA AAAAAAAAAA AAA 2003

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 11 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Ala Ser Leu Ala Asp Glu Tyr Glu Tyr Met Xaa Lys
 1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 9 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

Thr Glu Thr Ser Ser Ser Gly Leu Xaa Leu Lys
 1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12

- 97 -

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Ala Ser Leu Ala Asp Glu Tyr Glu Tyr Met Arg Lys
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 7 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Ala Gly Tyr Phe Ala Glu Xaa Ala Arg
1 5

- 98 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Thr Thr Glu Met Ala Ser Glu Gln Gly Ala
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

Ala Lys Glu Ala Leu Ala Ala Leu Lys
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Phe Val Leu Gln Ala Lys Lys
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Glu Thr Gln Pro Asp Pro Gly Gln Ile Leu Lys Lys Val Pro Met Val
1 5 10 15

- 99 -

Ile Gly Ala Tyr Thr
20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in positions 1, 3, 17 and 19 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

Xaa Glu Xaa Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys Lys Glu
1 5 10 15

Xaa Gly Xaa Gly Lys
20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

Ala Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 6 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

Lys Leu Glu Phe Leu Xaa Ala Lys
1 5

- 100 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

Xaa Val His Gln Val Trp Ala Ala Lys
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine, Xaa in position 11 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

Xaa Tyr Ile Phe Phe Met Glu Pro Glu Ala Xaa Ser Ser Gly
1 5 10

- 101 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine, Xaa in position 13 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

Xaa Leu Gly Ala Trp Gly Pro Pro Ala Phe Pro Val Xaa Tyr
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

Xaa Trp Phe Val Val Ile Glu Gly Lys
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

Xaa Ala Ser Pro Val Ser Val Gly Ser Val Gln Glu Leu Val Gln Arg
1 5 10 15

- 102 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

Xaa Val Cys Leu Leu Thr Val Ala Ala Leu Pro Pro Thr
 1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is Lysine or Arginine; Xaa in position 6 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

Xaa Asp Leu Leu Leu Xaa Val
 1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

Cys Thr Cys Gly Cys Cys Lys Cys Cys Arg Thr Thr Cys Ala Cys Arg
 1 5 10 15
 Cys Ala Gly Ala Ala Gly Gly Thr Cys Thr Thr Cys Thr Cys Cys Thr
 20 25 30
 Thr Cys Thr Cys Ala Gly Cys
 35

- 103 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

Cys Cys Thr Cys Gly Cys Thr Cys Cys Thr Thr Cys Thr Thr Cys Thr
1 5 10 15
Thr Gly Cys Cys Cys Thr Thr Cys
20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

Val His Gln Val Trp Ala Ala Lys
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 10 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

Tyr Ile Phe Phe Met Glu Pro Glu Ala Xaa Ser Ser Gly
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

- 104 -

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 12 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

Leu Gly Ala Trp Gly Pro Pro Ala Phe Pro Val Xaa Tyr
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

Trp Phe Val Val Ile Glu Gly Lys
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 46:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

Ala Ser Pro Val Ser Val Gly Ser Val Gln Glu Leu Val Gln Arg
1 5 10 15

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 47:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

Val Cys Leu Leu Thr Val Ala Ala Leu Pro Pro Thr
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9

- 105 -

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

Lys Val His Gln Val Trp Ala Ala Lys
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 12 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Xaa Lys
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 5 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

Asp Leu Leu Leu Xaa Val
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

TTYAARGGNG AYGCNCAYAC 20

- 106 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

CATRTAYTCR TAYTCRTCNG C 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

TGYTCNGANG CCATYTCNGT 20

- 107 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

TGYTCRCTNG CCATYTCNGT 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

CCDATNACCA TNGGNACYTT 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

GCNGCCCANNA CYTGRTGNAC 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

GCYTCNGGYT CCATRAARAA 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20

- 108 -

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

CCYTCDATNA CNACRAACCA 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

TCNGCRAART ANCCNGC 17

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 60:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

GCNGCNAGNG CYTCYTTNGC 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 61:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

GCNGCYAANG CYTCYTTNGC 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 62:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

- 109 -

TTYTTNGCYT GNAGNACRAA 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 63:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

TTYTTNGCYT GYAANACRAA 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 64:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

TGNACNAGYT CYTGNAC 17

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 65:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

TGNACYAAYT CYTGNAC 17

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 66:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

CATRTAYTCN CCNGARTCNG C 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 67:

- 110 -

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

CATRTAYTCN CCRCTRTCNG C 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

NGARTCNGCY AANGANGCYT T 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

NGARTCNGCN AGNGANGCYT T 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

RCTRTCNGCY AANGANGCYT T 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

- 111 -

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

RCTRTCNGCN AGNGANGCYT T 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 72:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

NGARTCNGCY AARCTNGCYT T 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 73:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

NGARTCNGCN AGRCTNGCYT T 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 74:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

RCTRTCNGCY AARCTNGCYT T 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 75:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

RCTRCTNGCN AGRCTNGCYT T 21

- 112 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 76:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

ACNACNGARA TGGCTCNNGA 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 77:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

ACNACNGARA TGGCAGYNGA 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 78:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

CAYCARGTNT GGGCNGCNAA 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 79:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

TTYGTNGTNA THGARGGNAA 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 80:

- 113 -

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

AARGGNGAYG CNCAYACNGA 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 81:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

GARGCNYTNG CNGCNYTNAA 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 82:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

GTNGGNTCNG TNCARGARYT 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 83:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

GTNGGNAGYG TNCARGARYT 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 84:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21
(B) TYPE: nucleic acid

- 114 -

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

NACYTTYTTN ARDATYTGNC C 21

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 85:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 417
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) **FEATURE:**

(D) OTHER INFORMATION: Xaa in positions 14, 23, 90, 100, 126, and 135 is a stop codon.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

TCTAA AAC TAC AGA GAC TGT ATT TTC ATG ATC ATC ATA GTT CTG TGA AAT AAG
Asn Tyr Arg Asp Cys Ile Phe Met Ile Ile Ile Val Leu Xaa Asn Ile

1			5			10			15							
CTT Leu	AAA Lys	CCG Pro	CTT Leu 20	TGG Trp	TCC Ser	TGA Xaa	TCT Ser	TGT Cys 25	AGG Arg	AAG Lys	TCA Ser	GAA Glu	CTT Leu 30	CGC Arg	ATT Ile	101
AGC Ser	AAA Lys	GCG Ala 35	TCA Ser	CTG Leu	GCT Ala	GAT Asp	TCT Ser 40	GGA Gly	GAA Glu	TAT Ser	ATG Met	TGC Cys 45	AAA Lys	GTG Val	ATC Ile	149
AGC Ser	AAA Lys 50	CTA Leu	GGA Gly	AAT Asn	GAC Asp	AGT Ser 55	GCC Ala	TCT Ser	GCC Ala	AAC Asn	ATC Ile 60	ACC Arg	ATT Ile	GTG Val	GAG Glu	197
TCA Ser 65	AAC Asn	GGT Gly	AAG Lys	AGA Arg	TGC Cys 70	CTA Leu	CTG Leu	CGT Arg	GCT Ala	ATT Ile 75	TCT Ser	CAG Gln	TCT Ser	CTA Leu 80	AGA Arg	245
GGA Gly	GTG Val	ATC Ile	AAG Lys 85	GTA Val	TGT Cys	GGT Gly	CAC His	ACT Thr	TGA Xaa 90	ATC Ile	ACG Thr	CAG Gln	GTG Val	TGT Cys 95	GAA Glu	293
ATC Ile	TCA Ser	TTG Cys	TGA Xaa 100	ACA Thr	AAT Asn	AAA Lys	AAT Asn 105	CAT His	GAA Glu	AGG Arg	AAA Lys	ACT Thr 110	CTA Leu	TGT Cys	TTG Leu	341
AAA Lys	TAT Tyr	CTT Leu 115	ATG Met	GGT Gly	CCT Pro	CCT Pro	GTA Val 120	AAG Lys	CTC Leu	TTC Phe	ACT Thr	CCA Pro 125	TAA Xaa	GGT Gly	GAA Glu	389
ATA Ile	GAC Asp 130	CTG Leu	AAA Lys	TAT Tyr	ATA Ile	TAG Xaa 135	ATT Ile	ATT Ile	T							417

- 115 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 86:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N at positions 19, 25, and 31
is Inosine. Y can be cytidine
or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

CCGAATTCTG CAGGARACNC ARCCNGAYCC NGG 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 87:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N at positions 14, 20, 23, 29,
and 35 is Inosine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

AAGGATCCTG CAGNGTRTAN GCNCCDATNA CCATNGG 37

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 88:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N at positions 16, 21, and 24
is Inosine. Y can be cytidine
or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

CCGAATTCTG CAGGCNGAYT CNGGNGARTA YATG 34

- 116 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: N at positions 16 and 25 is Inosine. Y can be cytidine or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

CCGAATTCTG CAGGCNGAYA GYGNGARTA YAT 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: N at positions 14, 15, 16, 26, and 29 is Inosine. Y can be cytidine or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

AAGGATCCTG CAGNNNCATR TAYTCNCCNG ARTC 34

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 91:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: N at positions 14, 15, 16, and 26 is Inosine. Y can be cytidine or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

AAGGATCCTG CAGNNNCATR TAYTCNCCRC TRTC 34

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 92:

- 117 -

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: N at positions 21, 28, and 31
is Inosine. Y can be cytidine
or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

CCGAATTCTG CAGCAYCARG TNTGGGCNGC NAA 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: N at position 31 is Inosine. Y
can be cytidine or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

CCGAATTCTG CAGATHTTYT TYATGGARCC NGARG 35

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: N at positions 18, 21, 24, 27,
and 33 is Inosine. Y can be
cytidine or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

CCGAATTCTG CAGGGGGNCC NCCNGCNTTY CCNGT 35

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 95:

(i) SEQUENCE CHARACTERISTICS:

- 118 -

(A) LENGTH: 33
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N at positions 21 and 24 is
Inosine. Y can be cytidine or
thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

CCGAATTCTG CAGTGGTTYG TNGTNATHGA RGG 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 96:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N at positions 17, 20, and 26
is Inosine. Y can be cytidine
or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96

AAGGATCCTG CAGYTTNGCN GCCCANACYT GRTG 34

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 97:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N at position 19 is Inosine. Y
can be cytidine or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

AAGGATCCTG CAGGCYTCNG GYTCCATRAA RAA 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 98:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33

- 119 -

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N at positions 16, 22, 25, 28,
and 31 is Inosine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

AAGGATCCTG CAGACNGGRA ANGCGGNGG NCC 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 99:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N at positions 17, 26, and 29
is Inosine. Y can be cytidine
or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

AAGGATCCTG CAGYTTNCCY TCDATNACNA CRAAC 35

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 100:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N at position 18 is Inosine. Y
can be cytidine or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

CATRTAYTCR TAYTCTCNGC AAGGATCCTG CAG 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 101:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 120 -

(ix) FEATURE:

- (D) OTHER INFORMATION: N at position 19, 25, and 31
is Inosine. Y can be cytidine
or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

CCGAATTCTG CAGAARGGNG AYGCNCAYAC NGA 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 102:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: N at position 3 and 18 is
Inosine. Y can be cytidine or
thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

GCNGCYAANG CYTCYTTNGC AAGGATCCTG CAG 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 103:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: N at position 3, 6, 9, and 18
is Inosine. Y can be cytidine
or thymidine.

- 121 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

GCNGCNAGNG CYTCYTTCG AAGGATCCTG CAG 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 104:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: N at position 3, 12, and 15 is Inosine. Y can be cytidine or thymidine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

TCNGCRAART ANCCNGCAAG GATCCTGCAG 30

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 105:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

CATCGATCTG CAGGCTGATT CTGGAGAATA TATGTGCA 38

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 106:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

AAGGATCCTG CAGCCACATC TCGAGTCGAC ATCGATT 37

- 122 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 107:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

CCGAATTCTG CAGTGATCAG CAAACTAGGA AATGACA 37

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 108:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

CATCGATCTG CAGCCTAGTT TGCTGATCAC TTTGCAC 37

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 109:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

AAGGATCCTG CAGTATATTC TCCAGAATCA GCCAGTG 37

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 110:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

AAGGATCCTG CAGGCACGCA GTAGGCATCT CTTA 34

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 111:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35

- 123 -

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

CCGAATTCTG CAGCAGAACT TCGCATTAGC AAAGC 35

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 112:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

CATCCCGGGA TGAAGAGTCA GGAGTCTGTG GCA 33

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 113:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

ATACCCGGGC TGCAGACAAT GAGATTTCAC ACACCTGCG 39

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 114:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

AAGGATCCTG CAGTTTGGA CCTGCCACAG ACTCCT 36

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 115:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

- 124 -

ATACCCGGGC TGCAGATGAG ATTTACACA CCTGCGTGA

39

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

His Gln Val Trp Ala Ala Lys Ala Ala Gly Leu Lys
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

Gly Gly Leu Lys Lys Asp Ser Leu Leu Thr Val Arg Leu Gly Ala Asn
1 5 10 15

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 12 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

Leu Gly Ala Trp Gly Pro Pro Ala Phe Pro Val Xaa Tyr
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

- 125 -

Leu Leu Thr Val Arg Leu Gly Ala Trp Gly His Pro Ala Phe Pro Ser
 1 5 10 15

Cys Gly Arg Leu Lys Glu Asp
 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 120:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 10 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

Tyr Ile Phe Phe Met Glu Pro Glu Ala Xaa Ser Ser Gly
 1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 121:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

Lys Glu Asp Ser Arg Tyr Ile Phe Phe Met Glu Pro Glu Ala Asn Ser
 1 5 10 15

Ser Gly Gly Pro Gly Arg Leu
 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 122:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

Val Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser
 1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 123:

(i) SEQUENCE CHARACTERISTICS:

- 126 -

(A) LENGTH: 16
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

Glu Tyr Lys Cys Leu Lys Phe Lys Trp Phe Lys Lys Ala Thr Val Met
1 5 10 15

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 124:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

Cys Glu Thr Ser Ser Glu Tyr Ser Ser Leu Lys Phe Lys Trp Phe Lys
1 5 10 15

Asn Gly Ser Glu Leu Ser Arg Lys Asn Lys
20 25

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 125:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in position 12 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Xaa Lys
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 126:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

Glu Leu Arg Ile Ser Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met
1 5 10 15

- 127 -

Glu Leu Arg Ile Ser Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met
 1 5 10 15
 Cys Lys Val Ile Ser Lys Leu
 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 127:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

Ala Ser Leu Ala Asp Glu Tyr Glu Tyr Met Arg Lys
 1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 128:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

Leu Arg Ile Ser Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Cys
 1 5 10 15
 Lys Val Ile Ser Lys Leu
 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 129:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 744
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

CCTGCAG CAT CAA GTG TGG GCG GCG AAA GCC GGG GGC TTG AAG AAG GAC TCG CTG
 55
 His Gln Val Trp Ala Ala Lys Ala Gly Gly Leu Lys Lys Asp Ser Leu
 1 5 10 15
 CTC ACC GTG CGC CTG GGC GCC TGG GGC CAC CCC GCC TTC CCC TCC TGC 103
 Leu Thr Val Arg Leu Gly Ala Trp Gly His Pro Ala Phe Pro Ser Cys
 20 25 30

- 128 -

GGG CGC CTC AAG GAG GAC AGC AGG TAC ATC TTC TTC ATG GAG CCC GAG	151
Gly Arg Leu Lys Glu Asp Ser Arg Tyr Ile Phe Phe Met Glu Pro Glu	
35 40 45	
GCC AAC AGC AGC GGC GGG CCC GGC CGC CTT CCG AGC CTC CTT CCC CCC	199
Ala Asn Ser Ser Gly Gly Pro Gly Arg Leu Pro Ser Leu Leu Pro Pro	
50 55 60	
TCT CGA GAC GGG CCG GAA CCT CAA GAA GGA GGT CAG CCG GGT GCT GTG	247
Ser Arg Asp Gly Pro Glu Pro Gln Glu Gly Gln Pro Gly Ala Val	
65 70 75 80	
CAA CGG TGC GCC TTG CCT CCC CGC TTG AAA GAG ATG AAG AGT CAG GAG	295
Gln Arg Cys Ala Leu Pro Pro Arg Leu Lys Glu Met Lys Ser Gln Glu	
85 90 95	
TCT GTG GCA GGT TCC AAA CTA GTG CTT CGG TGC GAG ACC AGT TCT GAA	343
Ser Val Ala Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu	
100 105 110	
TAC TCC TCT CTC AAG TTC AAG TGG TTC AAG AAT GGG AGT GAA TTA AGC	391
Tyr Ser Ser Leu Lys Phe Lys Trp Phe Lys Asn Gly Ser Glu Leu Ser	
115 120 125	
CGA AAG AAC AAA CCA GAA AAC ATC AAG ATA CAG AAA AGG CCG GGG AAG	439
Arg Lys Asn Lys Pro Glu Asn Ile Lys Ile Gln Lys Arg Pro Gly Lys	
130 135 140	
TCA GAA CTT CGC ATT AGC AAA GCG TCA CTG GCT GAT TCT GGA GAA TAT	487
Ser Glu Leu Arg Ile Ser Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr	
145 150 155 160	
ATG TGC AAA GTG ATC AGC AAA CTA GGA AAT GAC AGT GCC TCT GCC AAC	535
Met Cys Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala Asn	
165 170 175	
ATC ACC ATT GTG GAG TCA AAC GGT AAG AGA TGC CTA CTG CGT GCT ATT	583
Ile Thr Ile Val Glu Ser Asn Gly Lys Arg Cys Leu Leu Arg Ala Ile	
180 185 190	
TCT CAG TCT CTA AGA GGA GTG ATC AAG GTA TGT GGT CAC ACT	625
Ser Gln Ser Leu Arg Gly Val Ile Lys Val Cys Gly His Thr	
195 200 205	
TGAATCACGC AGGTGTGTGA AATCTCATTG TGAACAAATA AAAATCATGA AAGGAAAAAA	685
AAAAAAAAAA AATCGATGTC GACTCGAGAT GTGGCTGCAG GTCGACTCTA GAGGATCCC	744

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 130:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1193
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

- 129 -

CCTGCAG CAT CAA GTG TGG GCG GCG AAA GCC GGG GGC TTG AAG AAG GAC TCG CTG
 55
 His Gln Val Trp Ala Ala Lys Ala Gly Gly Leu Lys Lys Asp Ser Leu
 1 5 10 15

CTC ACC GTG CGC CTG GGC GCC TGG GGC CAC CCC GCC TTC CCC TCC TGC 103
 Leu Thr Val Arg Leu Gly Ala Trp Gly His Pro Ala Phe Pro Ser Cys
 20 25 30

GGG CGC CTC AAG GAG GAC AGC AGG TAC ATC TTC TTC ATG GAG CCC GAG 151
 Gly Arg Leu Lys Glu Asp Ser Arg Tyr Ile Phe Phe Met Glu Pro Glu
 35 40 45

GCC AAC AGC AGC GGC GGG CCC GGC CGC CTT CCG AGC CTC CTT CCC CCC 199
 Ala Lys Ser Ser Gly Gly Pro Gly Arg Leu Pro Ser Leu Leu Pro Pro
 50 55 60

TCT CGA GAC GGG CCG GAA CCT CAA GAA GGA GGT CAG CCG GGT GCT GTG 247
 Ser Arg Asp Gly Pro Glu Pro Gln Glu Gly Gly Gln Pro Gly Ala Val
 65 70 75 80

CAA CGG TGC GCC TTG CCT CCC CGC TTG AAA GAG ATG AAG AGT CAG GAG 295
 Gln Arg Cys Ala Leu Pro Pro Arg Leu Lys Glu Met Lys Ser Gln Glu
 85 90 95

TCT GTG GCA GGT TCC AAA CTA GTG CTT CGG TGC GAG ACC AGT TCT GAA 343
 Ser Val Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu
 100 105 110

TAC TCC TCT CTC AAG TTC AAG TGG TTC AAG AAT GGG AGT GAA TTA AGC 391
 Tyr Ser Ser Leu Lys Phe Lys Trp Phe Lys Asn Gly Ser Glu Leu Ser
 115 120 125

CGA AAG AAC AAA CCA GAA AAC ATC AAG ATA CAG AAA AGG CCG GGG AAG 439
 Arg Lys Asn Lys Gly Gly Asn Ile Lys Ile Gln Lys Arg Pro Gly Lys
 130 135 140

TCA GAA CTT CGC ATT AGC AAA GCG TCA CTG GCT GAT TCT GGA GAA TAT 487
 Ser Glu Leu Arg Ile Ser Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr
 145 150 155 160

ATG TGC AAA GTG ATC AGC AAA CTA GGA AAT GAC AGT GCC TCT GCC AAC 535
 Met Cys Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala Asn
 165 170 175

ATC ACC ATT GTG GAG TCA AAC GCC ACA TCC ACA TCT ACA GCT GGG ACA 583
 Ile Thr Ile Val Glu Ser Asn Ala Thr Ser Thr Ser Thr Ala Gly Thr
 180 185 190

AGC CAT CTT GTC AAG TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG AAT 631
 Ser His Leu Val Lys Ser Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
 195 200 205

GGA GGC GAG TGC TTC ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA TAC 679
 Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
 210 215 220

TTG TGC AAG TGC CAA CCT GGA TTC ACT GGA GCG AGA TGT ACT GAG AAT 727
 Leu Cys Lys Cys Gln Pro Gly Phe Thr Gly Ala Arg Cys Thr Glu Asn
 225 230 235 240

- 130 -

GTG CCC ATG AAA GTC CAA ACC CAA GAA AGT GCC CAA ATG AGT TTA CTG 775
 Val Pro Met Lys Val Gln Thr Gln Glu Ser Ala Gln Met Ser Leu Leu
 245 250 255

GTG ATC GCT GCC AAA ACT ACG TAATGGCCAG CTTCTACAGT ACGTCCACTC 826
 Val Ile Ala Ala Lys Thr Thr
 260

CCTTTCTGTC TCTGCCTGAA TAGCGCATCT CAGTCGGTGC CGCTTTCTTG TTGCCGCATC 886
 TCCCCTCAGA TTCCTCCTAG AGCTAGATGC GTTTTACCAG GTCTAACATT GACTGCCTCT 946
 GCCTGTGCGA TGAGAACATT AACACAAGCG ATTGTATGAC TTCCTCTGTC CGTGACTAGT 1006
 GGGCTCTGAG CTACTCGTAG GTGCGTAAGG CTCAGTGTT TCTGAAATTG ATCTGAATT 1066
 ACTGTGATAC GACATGATAG TCCCTCTCAC CCAGTGCAAT GACAATAAAG GCCTTGAAAA 1126
 GTCAAAAAAA AAAAAAATAA AAAAAATCGA TGTCGACTCG AGATGTGGCT GCAGGTGCGAC 1186
 TCTAGAG 1193

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 131:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1108
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

CCTGCAG CAT CAA GTG TGG GCG GCG AAA GCC GGG GGC TTG AAG AAG GAC TCG CTG 55
 His Gln Val Trp Ala Ala Lys Ala Gly Gly Leu Lys Lys Asp Ser Leu
 1 5 10 15

CTC ACC GTG CGC CTG GGC GCC TGG GGC CAC CCC GCC TTC CCC TCC TGC 103
 Leu Thr Val Arg Leu Gly Ala Trp Gly His Pro Ala Phe Pro Ser Cys
 20 25 30

GGG CGC CTC AAG GAG GAC AGC AGG TAC ATC TTC TTC ATG GAG CCC GAG 151
 Gly Arg Leu Lys Glu Asp Ser Arg Tyr Ile Phe Phe Met Glu Pro Glu
 35 40 45

GCC AAC AGC AGC GGC GGG CCC GGC CGC CTT CCG AGC CTC CTT CCC CCC 199
 Ala Asn Ser Ser Gly Gly Pro Gly Arg Leu Pro Ser Leu Leu Pro Pro
 50 55 60

TCT CGA GAC GGG CCG GAA CCT CAA GAA GGA GGT CAG CCG GGT GCT GTG 247
 Ser Arg Asp Gly Pro Glu Pro Gln Glu Gly Gly Gln Pro Gly Ala Val
 65 70 75 80

CAA CGG TGC GCC TTG CCT CCC CGC TTG AAA GAG ATG AAG AGT CAG GAG 295
 Gln Arg Cys Ala Leu Pro Pro Arg Leu Lys Glu Met Lys Ser Gln Glu
 85 90 95

TCT GTG GCA GGT TCC AAA CTA GTG CTT CGG TGC GAG ACC AGT TCT GAA 343
 Ser Val Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu

- 131 -

	100	105	110	
TAC TCC TCT CTC AAG TTC AAG TGG TTC AAG AAT GGG AGT GAA TTA AGC				391
Tyr Ser Ser Leu Lys Phe Lys Trp Phe Lys Asn Gly Ser Glu Leu Ser				
	115	120	125	
CGA AAG AAC AAA CCA GAA AAC ATC AAG ATA CAG AAA AGG CCG GGG AAG				439
Arg Lys Asn Lys Pro Glu Asn Ile Lys Ile Gln Lys Arg Pro Pro Lys				
	130	135	140	
TCA GAA CTT CGC ATT AGC AAA GCG TCA CTG GCT GAT TCT GGA GAA TAT				487
Ser Glu Leu Arg Ile Ser Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr				
	145	150	155	160
ATG TGC AAA GTG ATC AGC AAA CTA GGA AAT GAC AGT GCC TCT GCC AAC				535
Met Cys Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala Asn				
	165	170	175	
ATC ACC ATT GTG GAG TCA AAC GCC ACA TCC ACA TCT ACA GCT GGG ACA				583
Ile Arg Ile Val Glu Ser Asn Ala Thr Ser Thr Ser Thr Ala Gly Thr				
	180	185	190	
AGC CAT CTT GTC AAG TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG AAT				631
Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn				
	195	200	205	
GGA GGC GAG TGC TTC ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA TAC				679
Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr				
	210	215	220	
TTG TGC AAG TGC CCA AAT GAG TTT ACT GGT GAT CGC TGC CAA AAC TAC				727
Leu Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn Tyr				
	225	230	235	240
GTA ATG GCC AGC TTC TAC AGT ACG TCC ACT CCC TTT CTG TCT CTG CCT				775
Val Met Ala Ser Phe Tyr Ser Thr Ser Thr Pro Phe Leu Ser Leu Pro				
	245	250	255	
GAA TAGCGCATCT CAGTCGGTGC CGCTTTCTTG TTGCGGCATC TCCCCTCAGA TTCCGCCTAG				
838				
Glu				
AGCTAGATGC GTTTTACCAG GTCTAACATT GACTGCCTCT GCCTGTGCGA TGAGAACATT				898
AACACAAGCG ATTGTATGAC TTCCTCTGTC CGTGACTAGT GGGCTCTGAG CTACTCGTAG				958
GTGCGTAAGG CTCCAGTGTT TCTGAAATTG ATCTTGAATT ACTGTGATAC GACATGATAG				1018
TCCCTCTCAC CCAGTGCAAT GACAATAAAG GCCTTGAAAA GTCAAAAAAA AAAAAAAAAA				1078
AAAAATCGAT GTCGACTCGA GATGTGGCTG				1108

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 132:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 559
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

- 132 -

(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N in position 214 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

AGTTTCCCCC CCCAACTTGT CGGAACTCTG GGCTCGCGCG CAGGGCAGGA GCGGAGCGGC
60

GGCGGCTGCC CAGGCGATGC GAGCGCGGGC CGGACGGTAA TCGCCTCTCC CTCCTCGGGC
120

TGCGAGCGCG CCGGACCGAG GCAGCGACAG GAGCGGACCG CGGCGGGAAC CGAGGACTCC
180

CCAGCGGCGC GCCAGCAGGA GCCACCCCGC GAGNCGTGCG ACCGGGACGG AGCGCCCCGC
240

AGTCCCAGGT GGCCCGGACC GCACGTTGCG TCCCCGCGCT CCCC GCCGGC GACAGGAGAC
300

GCTCCCCCCC ACGCCGCGCG CGCCTCGGCC CGGTCGCTGG CCCGCCTCCA CTCGCGGGAC
360

AAACTTTTCC CGAAGCCGAT CCCAGCCCTC GGACCCAAAC TTGTCGCGCG TCGCCTTCGC
420

CGGGAGCCGT CCGCGCAGAG CGTGCACTTC TCGGGCGAG ATG TCG GAG CGC AGA
474

Met Ser Glu Arg Arg
1 5

GAA GGC AAA GGC AAG GGG AAG GGC GGC AAG AAG GAC CGA GGC TCC GGG
522

Glu Gly Lys Gly Lys Gly Lys Gly Gly Lys Lys Asp Arg Gly Ser Gly
10 15 20

AAG AAG CCC GTG CCC GCG GCT GGC GGC CCG AGC CCA G
559

Lys Lys Pro Val Pro Ala Ala Gly Gly Pro Ser Pro Ala
25 30

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 133:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 252
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N in position 8 could be either A or G.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

- 133 -

CC CAT CAN GTG TGG GCG GCG AAA GCC GGG GGC TTG AAG AAG GAC TCG
47

His Gln Val Trp Ala Ala Lys Ala Gly Gly Leu Lys Lys Asp Ser
1 5 10 15

CTG CTC ACC GTG CGC CTG GGC GCC TGG GGC CAC CCC GCC TTC CCC TCC
95

Leu Leu Thr Val Arg Leu Gly Ala Trp Gly His Pro Ala Phe Pro Ser
20 25 30

TGC GGG CGC CTC AAG GAG GAC AGC AGG TAC ATC TTC TTC ATG GAG CCC
143

Cys Gly Arg Leu Lys Glu Asp Ser Arg Tyr Ile Phe Phe Met Glu Pro
35 40 45

GAG GCC AAC AGC AGC GGC GGG CCC GGC CGC CTT CCG AGC CTC CTT CCC
191

Glu Ala Asn Ser Ser Gly Gly Pro Gly Arg Leu Pro Ser Leu Leu Pro
50 55 60

CCC TCT CGA GAC GGG CCG GAA CCT CAA GAA GGA GGT CAG CCG GGT GCT
239

Pro Ser Arg Asp Gly Pro Glu Pro Gln Glu Gly Gly Gln Pro Gly Ala
65 70 75

GTG CAA CGG TGC G

252

Val Gln Arg Cys

80

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 134:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 178
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

CCT TGC CTC CCC GCT TGA AAG AGA TGA AGA GTC AGG AGT CTG TGG CAG
48

Leu Pro Pro Arg Leu Lys Glu His Lys Ser Gln Glu Ser Val Ala Gly
1 5 10 15

GTT CCA AAC TAG TGC TTC GGT GCG AGA CCA GTT CTG AAT ACT CCT CTC
96

Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser Ser Leu
20 25 30

TCA AGT TCA AGT GGT TCA AGA ATG GGA GTG AAT TAA GCC GAA AGA ACA
144

Lys Phe Lys Trp Phe Lys Asn Gly Ser Glu Leu Ser Arg Lys Asn Lys
35 40 45

AAC CAC AAA ACA TCA AGA TAC AGA AAA GGC CGG G
178

Pro Gly Asn Ile Lys Ile Gln Lys Arg Pro Gly

- 134 -

50

55

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 135:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 122
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

G AAG TCA GAA CTT CGC ATT AGC AAA GCG TCA CTG GCT GAT TCT GGA
46
Lys Ser Glu Leu Arg Ile Ser Lys Ala Ser Leu Ala Asp Ser Gly
1 5 10 15
GAA TAT ATG TGC AAA GTG ATC AGC AAA CTA GGA AAT GAC AGT GCC TCT
94
Glu Tyr Met Cys Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser
20 25 30
GCC AAC ATC ACC ATT GTG GAG TCA AAC G
122
Ala Asn Ile Thr Ile Val Glu Ser Asn Ala
35

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 136:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 417
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

TCTAAACTA CAGAGACTGT ATTTTCATGA TCATCATAGT TCTGTGAAAT ATACTTAAAC
60
CGCTTTGGTC CTGATCTTGT AGG AAG TCA GAA CTT CGC ATT AGC AAA GCG
110
Lys Ser Glu Leu Arg Ile Ser Lys Ala
1 5
TCA CTG GCT GAT TCT GGA GAA TAT ATG TGC AAA GTG ATC AGC AAA CTA
158
Ser Leu Ala Asp Ser Gly Glu Tyr Met Cys Lys Val Ile Ser Lys Leu
10 15 20 25
GGA AAT GAC AGT GCC TCT GCC AAC ATC ACC ATT GTG GAG TCA AAC GGT
206
Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr Ile Val Glu Ser Asn Gly
30 35 40
AAG AGA TGC CTA CTG CGT GCT ATT TCT CAG TCT CTA AGA GGA GTG ATC
254

- 135 -

Lys Arg Cys Leu Arg Ala Ile Ser Gln Ser Leu Arg Gly Val Ile
 45 50 55

AAG GTA TGT GGT CAC ACT TGAATCACGC AGGTGTGTGA AATCTCATTG
 302

Lys Val Cys Gly His Thr
 60

TGAACAAATA AAAATCATGA AAGGAAACT CTATGTTTGA AATATCTTAT GGGTCCTCCT
 362

GTAAAGCTCT TCACTCCATA AGGTGAAATA GACCTGAAAT ATATATAGAT TATTT
 417

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 102
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

AG ATC ACC ACT GGC ATG CCA GCC TCA ACT GAG ACA GCG TAT GTG TCT
 47
 Glu Ile Thr Thr Gly Met Pro Ala Ser Thr Glu Thr Ala Tyr Val Ser
 1 5 10 15

TCA GAG TCT CCC ATT AGA ATA TCA GTA TCA ACA GAA GGA ACA AAT ACT
 95
 Ser Glu Ser Pro Ile Arg Ile Ser Val Ser Thr Glu Gly Thr Asn Thr
 20 25 30

TCT TCA T
 102
 Ser Ser Ser
 35

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 138:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

AAG TGC CAA CCT GGA TTC ACT GGA GCG AGA TGT ACT GAG AAT GTG CCC
 48
 Lys Cys Gln Pro Gly Phe Thr Gly Ala Arg Cys Thr Glu Asn Val Pro
 1 5 10 15

ATG AAA GTC CAA ACC CAA GAA
 69
 Met Lys Val Gln Thr Gln Glu
 20

- 136 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 139:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

AAG TGC CCA AAT GAG TTT ACT GGT GAT CGC TGC CAA AAC TAC GTA ATG
48
Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn Tyr Val Met
1 5 10 15

GCC AGC TTC TAC
60
Ala Ser Phe Tyr
20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 140:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

AGT ACG TCC ACT CCC TTT CTG TCT CTG CCT GAA TAG
36
Ser Thr Ser Thr Pro Phe Leu Ser Leu Pro Glu
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 141:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

AAG CAT CTT GGG ATT GAA TTT ATG GAG
27
Lys His Leu Gly Ile Glu Phe Met Glu
1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 142:

(i) SEQUENCE CHARACTERISTICS:

- 137 -

(A) LENGTH: 569
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

AAA GCG GAG GAG CTC TAC CAG AAG AGA GTG CTC ACC ATT ACC GGC ATT
 48
 Lys Ala Glu Glu Leu Tyr Gln Lys Arg Val Leu Thr Ile Thr Gly Ile
 1 5 10 15

TGC ATC GCG CTG CTC GTG GTT GGC ATC ATG TGT GTG GTG GTC TAC TGC
 96
 Cys Ile Ala Leu Leu Val Val Gly Ile Met Cys Val Val Val Tyr Cys
 20 25 30

AAA ACC AAG AAA CAA CGG AAA AAG CTT CAT GAC CGG CTT CGG CAG AGC
 144
 Lys Thr Lys Lys Gln Arg Lys Lys Leu His Asp Arg Leu Arg Gln Ser
 35 40 45

CTT CGG TCT GAA AGA AAC ACC ATG ATG AAC GTA GCC AAC GGG CCC CAC
 192
 Leu Arg Ser Glu Arg Asn Thr Met Met Asn Val Ala Asn Gly Pro His
 50 55 60

CAC CCC AAT CCG CCC CCC GAG AAC GTG CAG CTG GTG AAT CAA TAC GTA
 240
 His Pro Asn Pro Pro Pro Glu Asn Val Gln Leu Val Asn Gln Tyr Val
 65 70 75 80

TCT AAA AAT GTC ATC TCT AGC GAG CAT ATT GTT GAG AGA GAG GCG GAG
 288
 Ser Lys Asn Val Ile Ser Ser Glu His Ile Val Glu Arg Glu Ala Glu
 85 90 95

AGC TCT TTT TCC ACC AGT CAC TAC ACT TCG ACA GCT CAT CAT TCC ACT
 336
 Ser Ser Phe Ser Thr Ser His Tyr Thr Ser Thr Ala His His Ser Thr
 100 105 110

ACT GTC ACT CAG ACT CCC AGT CAC AGC TGG AGC AAT GGA CAC ACT GAA
 384
 Thr Val Thr Gln Thr Pro Ser His Ser Trp Ser Asn Gly His Thr Glu
 115 120 125

AGC ATC ATT TCG GAA AGC CAC TCT GTC ATC GTG ATG TCA TCC GTA GAA
 432
 Ser Ile Ile Ser Glu Ser His Ser Val Ile Val Met Ser Ser Val Glu
 130 135 140

AAC AGT AGG CAC AGC AGC CCG ACT GGG GGC CCG AGA GGA CGT CTC AAT
 480
 Asn Ser Arg His Ser Ser Pro Thr Gly Gly Pro Arg Gly Arg Leu Asn
 145 150 155 160

GGC TTG GGA GGC CCT CGT GAA TGT AAC AGC TTC CTC AGG CAT GCC AGA
 528
 Gly Leu Gly Gly Pro Arg Glu Cys Asn Ser Phe Leu Arg His Ala Arg

- 138 -

165 170 175
 GAA ACC CCT GAC TCC TAC CGA GAC TCT CCT CAT AGT G AAAG
 569
 Glu Thr Pro Asp Ser Tyr Arg Asp Ser Pro His Ser
 180 185

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 143:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 735
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

G TAT GTA TCA GCA ATG ACC ACC CCG GCT CGT ATG TCA CCT GTA GAT
 46
 Tyr Val Ser Ala Met Thr Thr Pro Ala Arg Met Ser Pro Val Asp
 1 5 10 15

TTC CAC ACG CCA AGC TCC CCC AAG TCA CCC CCT TCG GAA ATG TCC CCG
 94
 Phe His Thr Pro Ser Ser Pro Lys Ser Pro Pro Ser Glu Met Ser Pro
 20 25 30

CCC GTG TCC AGC ACG ACG GTC TCC ATG CCC TCC ATG GCG GTC AGT CCC
 142
 Pro Val Ser Ser Thr Thr Val Ser Met Pro Ser Met Ala Val Ser Pro
 35 40 45

TTC GTG GAA GAG GAG AGA CCC CTG CTC CTT GTG ACG CCA CCA CGG CTG
 190
 Phe Val Glu Glu Glu Arg Pro Leu Leu Leu Val Thr Pro Pro Arg Leu
 50 55 60

CGG GAG AAG TAT GAC CAC CAC GCC CAG CAA TTC AAC TCG TTC CAC TGC
 238
 Arg Glu Lys Tyr Asp His His Ala Gln Gln Phe Asn Ser Phe His Cys
 65 70 75

AAC CCC GCG CAT GAG AGC AAC AGC CTG CCC CCC AGC CCC TTG AGG ATA
 286
 Asn Pro Ala His Glu Ser Asn Ser Leu Pro Pro Ser Pro Leu Arg Ile
 80 85 90 95

GTG GAG GAT GAG GAA TAT GAA ACG ACC CAG GAG TAC GAA CCA GCT CAA
 334
 Val Glu Asp Glu Glu Tyr Glu Thr Thr Gln Glu Tyr Glu Pro Ala Gln
 100 105 110

GAG CCG GTT AAG AAA CTC ACC AAC AGC AGC CCG CCG GCC AAA AGA ACC
 382
 Glu Pro Val Lys Lys Leu Thr Asn Ser Ser Arg Arg Ala Lys Arg Thr
 115 120 125

AAG CCC AAT GGT CAC ATT GCC CAC AGG TTG GAA ATG GAC AAC AAC ACA
 430

- 139 -

Lys Pro Asn Gly His Ile Ala His Arg Leu Glu Met Asp Asn Asn Thr
 130 135 140
 GGC GCT GAC AGC AGT AAC TCA GAG AGC GAA ACA GAG GAT GAA AGA GTA
 478
 Gly Ala Asp Ser Ser Asn Ser Glu Ser Glu Thr Glu Asp Glu Arg Val
 145 150 155
 GGA GAA GAT ACG CCT TTC CTG GCC ATA CAG AAC CCC CTG GCA GCC AGT
 526
 Gly Glu Asp Thr Pro Phe Leu Ala Ile Gln Asn Pro Leu Ala Ala Ser
 160 165 170 175
 CTC GAG GCG GCC CCT GCC TTC CGC CTG GTC GAC AGC AGG ACT AAC CCA
 574
 Leu Glu Ala Ala Pro Ala Phe Arg Leu Val Asp Ser Arg Thr Asn Pro
 180 185 190
 ACA GGC GGC TTC TCT CCG CAG GAA GAA TTG CAG GCC AGG CTC TCC GGT
 622
 Thr Gly Gly Phe Ser Pro Gln Glu Glu Leu Gln Ala Arg Leu Ser Gly
 195 200 205
 GTA ATC GCT AAC CAA GAC CCT ATC GCT GTC TAAAACCGAA ATACACCCAT
 672
 Val Ile Ala Asn Gln Asp Pro Ile Ala Val
 210 215
 AGATTCACCT GTAAACTTT ATTTTATATA ATAAAGTATT CCACCTTAAA TTAAACAA
 730

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1654
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

AGTTTCCCCC CCCAACTTGT CGGAACTCTG GGCTCGCGCG CAGGGCAGGA GCGGAGCGGC
 60
 GCGCGCTGCC CAGGCGATGC GAGCGCGGGC CGGACGGTAA TCGCCTCTCC CTCCTCGGGC
 120
 TGCGAGCGCG CCGGACCGAG GCAGCGACAG GAGCGGACCG CGGCGGGAAC CGAGGACTCC
 180
 CCAGCGGCGC GCCAGCAGGA GCCACCCCGC GAGCGTGCGA CCGGGACGGA GCGCCCGCCA
 240
 GTCCCAGGTG GCGCGGACCG CACGTTGCGT CCGCGCGCTC CCGCGCGGCG ACAGGAGACG
 300
 CTCCTCCCCA CGCGCGCGCG GCCTCGGCCC GGTGCTGGC CCGCCTCCAC TCGGGGACA
 360

- 140 -

AAC TTTTCCC GAAGCCGATC CCAGCCCTCG GACCCAAACT TGTCGCGCGT CGCCTTCGCC
420

GGGAGCCGTC CGCGCAGAGC GTGCACTTCT CGGGCGAG ATG TCG GAG CGC AGA
475

Met Ser Glu Arg Arg
1 5

GAA GGC AAA GGC AAG GGG AAG GGC GGC AAG AAG GAC CGA GGC TCC GGG
523
Glu Gly Lys Gly Lys Gly Lys Gly Gly Lys Lys Asp Arg Gly Ser Gly
10 15 20

AAG AAG CCC GTG CCC GCG GCT GGC GGC CCG AGC CCA GCC TTG CCT CCC
571
Lys Lys Pro Val Pro Ala Ala Gly Gly Pro Ser Pro Ala Leu Pro Pro
25 30 35

CGC TTG AAA GAG ATG AAG ATG CAG GAG TCT GTG GCA GGT TCC AAA CTA
Arg Leu Lys Glu Met Lys Ser Gln Glu Ser Val Ala Gly Ser Lys Leu
40 45 50

619

GTG CTT CGG TGC GAG ACC AGT TCT GAA TAC TCC TCT CTC AAG TTC AAG
667
Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser Ser Leu Lys Phe Lys
55 60 65

TGG TTC AAG AAT GGG AGT GAA TTA AGC CGA AAG AAC AAA CCA CAA AAC
715
Trp Phe Lys Asn Gly Ser Glu Leu Ser Arg Lys Asn Lys Pro Gln Asn
70 75 80 85

ATC AAG ATA CAG AAA AGG CCG GGG AAG TCA GAA CTT CGC ATT AGC AAA
763
Ile Lys Ile Gln Lys Arg Pro Gly Lys Ser Glu Leu Arg Ile Ser Lys
90 95 100

GCG TCA CTG GCT GAT TCT GGA GAA TAT ATG TGC AAA GTG ATC AGC AAA
811
Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Cys Lys Val Ile Ser Lys
105 110 115

CTA GGA AAT GAC AGT GCC TCT GCC AAC ATC ACC ATT GTG GAG TCA AAC
859
Leu Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr Ile Val Glu Ser Asn
120 125 130

GAG ATC ACC ACT GGC ATG CCA GCC TCA ACT GAG ACA GCG TAT GTG TCT
907
Glu Ile Thr Thr Gly Met Pro Ala Ser Thr Glu Thr Ala Tyr Val Ser
135 140 145

TCA GAG TCT CCC ATT AGA ATA TCA GTA TCA ACA GAA GGA ACA AAT ACT
955
Ser Glu Ser Pro Ile Arg Ile Ser Val Ser Thr Glu Gly Thr Asn Thr
150 155 160 165

TCT TCA TCC ACA TCC ACA TCT ACA GCT GGG ACA AGC CAT CTT GTC AAG
1003
Ser Ser Ser Thr Ser Thr Ser Thr Ala Gly Thr Ser His Leu Val Lys
170 175 180

- 141 -

TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG AAT GGA GGC GAG TGC TTC
1051
Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys Phe
185 190 195

ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA TAC TTG TGC AAG TGC CCA
 1099
 Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu Cys Lys Cys Pro
 200 205 210

AAT GAG TTT ACT GGT GAT CGC TGC CAA AAC TAC GTA ATG GCC AGC TTC
 1147
 Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn Tyr Val Met Ala Ser Phe
 215 220 225

TAC AGT ACG TCC ACT CCC TTT CTG TCT CTG CCT GAA TAGGCGCATG
1193
Tyr Ser Thr Ser Thr Pro Phe Leu Ser Leu Pro Glu
230 235 240

CTCAGTCGGT GCCGCTTTCT TGTGCGCA TCTCCCCTCA GATTCAACCT AGAGCTAGAT
1253

CGGTTTTACC AGGTCTAACA TTGACTGCCT CTGCCTGTCG CATGAGAACA TTAACACAAG
1313

CGATTGTATG ACTTCCTCTG TCCGTGACTA GTGGGCTCTG AGCTACTCGT AGGTGCGTAA
1373

GGCTCCAGTG TTTCTGAAAT TGATCTTGAA TTA CTGTGAT ACGACATGAT AGTCCCTCTC
1433

ACCCAGTGCA ATGACAATAA AGGCCTTGAA AAGTCTCACT TTTATTGAGA AAATAAAAAAT
1493

CGTCCACGG GACAGTCCCT CTTCTTTATA AAATGACCCT ATCCTTGAAA AGGAGGTGTG
1553

TTAAGTTGTA ACCAGTACAC ACTTGAAATG ATGGTAAGTT CGCTTCGGTT CAGAATGTGT
1613

TCTTTCTGAC AAATAAACAG AATAAAAAA AAAAAAAAAA A
1654

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 145:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1140
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

CAT CAN GTG TGG GCG GCG AAA GCC GGG GGC TTG AAG AAG GAC TCG CTG
48
His Gln Val Trp Ala Ala Lys Ala Gly Gly Leu Lys Lys Asp Ser Leu
1 5 10 15

- 142 -

CTC ACC GTG CGC CTG GGC GCC TGG GGC CAC CCC GCC TTC CCC TCC TGC
 96
 Leu Thr Val Arg Leu Gly Ala Trp Gly His Pro Ala Phe Pro Ser Cys
 20 25 30

GGG CGC CTC AAG GAG GAC AGC AGG TAC ATC TTC TTC ATG GAG CCC GAG
 144
 Gly Arg Leu Lys Glu Asp Ser Arg Tyr Ile Phe Phe Met Glu Pro Glu
 35 40 45

GCC AAC AGC AGC GGC GGG CCC GGC CGC CTT CCG AGC CTC CTT CCC CCC
 192
 Ala Asn Ser Ser Gly Gly Pro Gly Arg Leu Pro Ser Leu Leu Pro Pro
 50 55 60

TCT CGA GAC GGG CCG GAA CCT CAA GAA GGA GGT CAG CCG GGT GCT GTG
 240
 Ser Arg Asp Gly Pro Glu Pro Gln Glu Gly Gly Gln Pro Gly Ala Val
 65 70 75 80

CAA CGG TGC GCC TTG CCT CCC CGC TTG AAA GAG ATG AAG AGT CAG GAG
 288
 Gln Arg Cys Ala Leu Pro Pro Arg Leu Lys Glu Met Lys Ser Gln Glu
 85 90 95

TCT GTG GCA GGT TCC AAA CTA GTG CTT CGG TGC GAG ACC AGT TCT GAA
 336
 Ser Val Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu
 100 105 110

TAC TCC TCT CTC AAG TTC AAG TGG TTC AAG AAT GGG AGT GAA TTA AGC
 384
 Tyr Ser Ser Leu Lys Phe Lys Trp Phe Lys Asn Gly Ser Glu Leu Ser
 115 120 125

CGA AAG AAC AAA CCA GAA AAC ATC AAG ATA CAG AAA AGG CCG GGG AAG
 432
 Arg Lys Asn Lys Pro Glu Asn Ile Lys Ile Gln Lys Arg Pro Gly Lys
 130 135 140

TCA GAA CTT CGC ATT AGC AAA GCG TCA CTG GCT GAT TCT GGA GAA TAT
 480
 Ser Glu Leu Arg Ile Ser Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr
 145 150 155 160

ATG TGC AAA GTG ATC AGC AAA CTA GGA AAT GAC AGT GCC TCT GCC AAC
 528
 Met Cys Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala Asn
 165 170 175

ATC ACC ATT GTG GAG TCA AAC GCC ACA TCC ACA TCT ACA GCT GGG ACA
 576
 Ile Thr Ile Val Glu Ser Asn Ala Thr Ser Thr Ser Thr Ala Gly Thr
 180 185 190

AGC CAT CTT GTC AAG TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG AAT
 624
 Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
 195 200 205

- 143 -

GGA GGC GAG TGC TTC ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA TAC
 672
 Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
 210 215 220
 TTG TGC AAG TGC CAA CCT GGA TTC ACT GGA GCG AGA TGT ACT GAG AAT
 720
 Leu Cys Lys Cys Gln Pro Gly Phe Thr Gly Ala Arg Cys Thr Glu Asn
 225 230 235 240
 GTG CCC ATG AAA GTC CAA ACC CAA GAA AAG TGC CCA AAT GAG TTT ACT
 768
 Val Pro Met Lys Val Gln Thr Gln Glu Lys Cys Pro Asn Glu Phe Thr
 245 250 255
 GGT GAT CGC TGC CAA AAC TAC GTA ATG GCC AGC TTC TAC AGT ACG TCC
 816
 Gly Asp Arg Cys Gln Asn Tyr Val Met Ala Ser Phe Tyr Ser Thr Ser
 260 265 270
 ACT CCC TTT CTG TCT CTG CCT GAA TAGCGCATCT CAGTCGGTGC CGCTTTCTTG
 870
 Thr Pro Phe Leu Ser Leu Pro Glu
 275 280
 TTGCCGCATC TCCCCTCAGA TTCNCCTAG AGCTAGATGC GTTTTACCAG GTCTAACATT
 930
 GACTGCCTCT GCCTGTCGCA TGAGAACATT AACACAAGCG ATTGTATGAC TTCCTCTGTC
 990
 CGTGACTAGT GGGCTCTGAG CTACTCGTAG GTGCGTAAGG CTCCAGTGTT TCTGAAATTG
 1050
 ATCTTGAATT ACTGTGATAC GACATGATAG TCCCTCTCAC CCACTGCAAT GACAATAAAG
 1110
 GCCTTGAAAA GTCAAAAAAA AAAAAAAAAA
 1140

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 146:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1764
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

G AAG TCA GAA CTT CGC ATT AGC AAA GCG TCA CTG GCT GAT TCT GGA GAA
 49
 Lys Ser Glu Leu Arg Ile Ser Lys Ala Ser Leu Ala Asp Ser Gly Glu
 1 5 10 15
 TAT ATG TGC AAA GTG ATC AGC AAA CTA GGA AAT GAC AGT GCC TCT GCC
 97
 Tyr Met Cys Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala
 20 25 30

- 144 -

AAC ATC ACC ATT GTG GAG TCA AAC GCC ACA TCC ACA TCT ACA GCT GGG
 145
 Asn Ile Thr Ile Val Glu Ser Asn Ala Thr Ser Thr Ser Thr Ala Gly
 35 40 45

ACA AGC CAT CTT GTC AAG TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG
 193
 Thr Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val
 50 55 60

AAT GGA GGC GAC TGC TTC ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA
 241
 Asn Gly Gly Asp Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg
 65 70 75 80

TAC TTG TGC AAG TGC CAA CCT GGA TTC ACT GGA GCG AGA TGT ACT GAG
 289
 Tyr Leu Cys Lys Cys Gln Pro Gly Phe Thr Gly Ala Arg Cys Thr Glu
 85 90 95

AAT GTG CCC ATG AAA GTC CAA ACC CAA GAA AAA GCG GAG GAG CTC TAC
 337
 Asn Val Pro Met Lys Val Gln Thr Gln Glu Lys Ala Glu Glu Leu Tyr
 100 105 110

CAG AAG AGA GTG CTC ACC ATT ACC GGC ATT TGC ATC GCG CTG CTC GTG
 385
 Gln Lys Arg Val Leu Thr Ile Thr Gly Ile Cys Ile Ala Leu Leu Val
 115 120 125

GTT GGC ATC ATG TGT GTG GTG GTC TAC TGC AAA ACC AAG AAA CAA CGG
 433
 Val Gly Ile Met Cys Val Val Val Tyr Cys Lys Thr Lys Lys Gln Arg
 130 135 140

AAA AAG CTT CAT GAC CGG CTT CGG CAG AGC CTT CGG TCT GAA AGA AAC
 481
 Lys Lys Leu His Asp Arg Leu Arg Gln Ser Leu Arg Ser Glu Arg Asn
 145 150 155 160

ACC ATG ATG AAC GTA GCC AAC GGG CCC CAC CAC CCC AAT CCG CCC CCC
 529
 Thr Met Met Asn Val Ala Asn Gly Pro His His Pro Asn Pro Pro Pro
 165 170 175

GAG AAC GTG CAG CTG GTG AAT CAA TAC GTA TCT AAA AAT GTC ATC TCT
 577
 Glu Asn Val Gln Leu Val Asn Gln Tyr Val Ser Lys Asn Val Ile Ser
 180 185 190

AGC GAG CAT ATT GTT GAG AGA GAG GCG GAG AGC TCT TTT TCC ACC AGT
 625
 Ser Glu His Ile Val Glu Arg Glu Ala Glu Ser Ser Phe Ser Thr Ser
 195 200 205

CAC TAC ACT TCG ACA GCT CAT CAT TCC ACT ACT GTC ACT CAG ACT CCC
 673
 His Tyr Thr Ser Thr Ala His His Ser Thr Thr Val Thr Gln Thr Pro
 210 215 220

- 145 -

AGT CAC AGC TGG AGC AAT GGA CAC ACT GAA AGC ATC ATT TCG GAA AGC
 721
 Ser His Ser Trp Ser Asn Gly His Thr Glu Ser Ile Ile Ser Glu Ser
 225 230 235 240

CAC TCT GTC ATC GTG ATG TCA TCC GTA GAA AAC AGT AGG CAC AGC AGC
 769
 His Ser Val Ile Val Met Ser Ser Val Glu Asn Ser Arg His Ser Ser
 245 250 255

CCG ACT GGG GGC CCG AGA GGA CGT CTC AAT GGC TTG GGA GGC CCT CGT
 817
 Pro Thr Gly Gly Pro Arg Gly Arg Leu Asn Gly Leu Gly Gly Pro Arg
 260 265 270

GAA TGT AAC AGC TTC CTC AGG CAT GCC AGA GAA ACC CCT GAC TCC TAC
 865
 Glu Cys Asn Ser Phe Leu Arg His Ala Arg Glu Thr Pro Asp Ser Tyr
 275 280 285

CGA GAC TCT CCT CAT AGT GAA AGA CAT AAC CTT ATA GCT GAG CTA AGG
 913
 Arg Asp Ser Pro His Ser Glu Arg His Asn Leu Ile Ala Glu Leu Arg
 290 295 300

AGA AAC AAG GCC CAC AGA TCC AAA TGC ATG CAG ATC CAG CTT TCC GCA
 961
 Arg Asn Lys Ala His Arg Ser Lys Cys Met Gln Ile Gln Leu Ser Ala
 305 310 315 320

ACT CAT CTT AGA GCT TCT TCC ATT CCC CAT TGG GCT TCA TTC TCT AAG
 1009
 Thr His Leu Arg Ala Ser Ser Ile Pro His Trp Ala Ser Phe Ser Lys
 325 330 335

ACC CCT TGG CCT TTA GGA AGG TAT GTA TCA GCA ATG ACC ACC CCG GCT
 1057
 Thr Pro Trp Pro Leu Gly Arg Tyr Val Ser Ala Met Thr Thr Pro Ala
 340 345 350

CGT ATG TCA CCT GTA GAT TTC CAC ACG CCA AGC TCC CCC AAG TCA CCC
 1105
 Arg Met Ser Pro Val Asp Phe His Thr Pro Ser Ser Pro Lys Ser Pro
 355 360 365

CCT TCG GAA ATG TCC CCG CCC GTG TCC AGC ACG ACG GTC TCC ATG CCC
 1153
 Pro Ser Glu Met Ser Pro Pro Val Ser Ser Thr Thr Val Ser Met Pro
 370 375 380

TCC ATG GCG GTC AGT CCC TTC GTG GAA GAG GAG AGA CCC CTG CTC CTT
 1201
 Ser Met Ala Val Ser Pro Phe Val Glu Glu Glu Arg Pro Leu Leu Leu
 385 390 395 400

GTG ACG CCA CCA CGG CTG CGG GAG AAG TAT GAC CAC CAC GCC CAG CAA
 1249
 Val Thr Pro Pro Arg Leu Arg Glu Lys Tyr Asp His His Ala Gln Gln
 405 410 415

- 146 -

TTC AAC TCG TTC CAC TGC AAC CCC GCG CAT GAG AGC AAC AGC CTG CCC
 1297
 Phe Asn Ser Phe His Cys Asn Pro Ala His Glu Ser Asn Ser Leu Pro
 420 425 430

CCC AGC CCC TTG AGG ATA GTG GAG GAT GAG GAA TAT GAA ACG ACC CAG
 1345
 Pro Ser Pro Leu Arg Ile Val Glu Asp Glu Glu Tyr Glu Thr Thr Gln
 435 440 445

GAG TAC GAA CCA GCT CAA GAG CCG GTT AAG AAA CTC ACC AAC AGC AGC
 1393
 Glu Tyr Glu Pro Ala Gln Glu Pro Val Lys Lys Leu Thr Asn Ser Ser
 450 455 460

CGG CGG GCC AAA AGA ACC AAG CCC AAT GGT CAC ATT GCC CAC AGG TTG
 1441
 Arg Arg Ala Lys Arg Thr Lys Pro Asn Gly His Ile Ala His Arg Leu
 465 470 475 480

GAA ATG GAC AAC AAC ACA GGC GCT GAC AGC AGT AAC TCA GAG AGC GAA
 1489
 Glu Met Asp Asn Asn Thr Gly Ala Asp Ser Ser Asn Ser Glu Ser Glu
 485 490 495

ACA GAG GAT GAA AGA GTA GGA GAA GAT ACG CCT TTC CTG GCC ATA CAG
 1537
 Thr Glu Asp Glu Arg Val Gly Glu Asp Thr Pro Phe Leu Ala Ile Gln
 500 505 510

AAC CCC CTG GCA GCC AGT CTC GAG GCG GCC CCT GCC TTC CGC CTG GTC
 1585
 Asn Pro Leu Ala Ala Ser Leu Glu Ala Ala Pro Ala Phe Arg Leu Val
 515 520 525

GAC AGC AGG ACT AAC CCA ACA GGC GGC TTC TCT CCG CAG GAA GAA TTG
 1633
 Asp Ser Arg Thr Asn Pro Thr Gly Gly Phe Ser Pro Gln Glu Glu Leu
 530 535 540

CAG GCC AGG CTC TCC GGT GTA ATC GCT AAC CAA GAC CCT ATC GCT GTC
 1681
 Gln Ala Arg Leu Ser Gly Val Ile Ala Asn Gln Asp Pro Ile Ala Val
 545 550 555 560

TAAACCGAA ATACACCCAT AGATTCACCT GTAAACTTT ATTTTATATA ATAAAGTATT
 1741

CCACCTTAAA TTAAACAAAA AAA
 1764

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

- 147 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys
 1 5 10 15
 Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu Cys Lys Cys
 20 25 30
 Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn Tyr Val Met Ala Ser
 35 40 45
 Phe Tyr
 50

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys
 1 5 10 15
 Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu Cys Lys Cys
 20 25 30
 Gln Pro Gly Phe Thr Gly Ala Arg Cys Thr Glu Asn Val Pro Met Lys
 35 40 45
 Val Gln
 50

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

Glu Cys Leu Arg Lys Tyr Lys Asp Phe Cys Ile His Gly Glu Cys Lys
 1 5 10 15
 Tyr Val Lys Glu Leu Arg Ala Pro Ser Cys Lys Cys Gln Gln Glu Tyr
 20 25 30
 Phe Gly Glu Arg Cys Gly Glu Lys Ser Asn Lys Thr His Ser
 35 40 45

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: -150:

- 148 -

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 198
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

```

AGC CAT CTT GTC AAG TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG AAT
 48
Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
 1           5           10          15

GGA GGC GAG TGC TTC ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA TAC
 96
Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
          20          25          30

TTG TGC AAG TGC CCA AAT GAG TTT ACT GGT GAT CGC TGC CAA AAC TAC
144
Leu Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn Tyr
          35          40          45

GTA ATG GCC AGC TTC TAC AGT ACG TCC ACT CCC TTT CTG TCT CTG CCT
192
Val Met Ala Ser Phe Tyr Ser Thr Ser Thr Pro Phe Leu Ser Leu Pro
 50          55          60

GAA TAG
Glu
65

```

198

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 151:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

```

AGC CAT CTT GTC AAG TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG AAT
 48
Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
 1           5           10          15

GGA GGC GAG TGC TTC ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA TAC
 96
Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
          20          25          30

TTG TGC AAG TGC CAA CCT GGA TTC ACT GGA GCG AGA TGT ACT GAG AAT
144
Leu Cys Lys Cys Gln Pro Gly Phe Thr Gly Ala Arg Cys Thr Glu Asn
          35          40          45

```

- 149 -

GTG CCC ATG AAA GTC CAA ACC CAA GAA AAA GCG GAG GAG CTC TAC TAA
 192
 Val Pro Met Lys Val Gln Thr Gln Glu Lys Ala Glu Glu Leu Tyr
 50 55 60

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 152:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 183
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

AGC CAT CTT GTC AAG TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG AAT
 48
 Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
 1 5 10 15
 GGA GGC GAG TGC TTC ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA TAC
 96
 Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
 20 25 30
 TTG TGC AAG TGC CCA AAT GAG TTT ACT GGT GAT CGC TGC CAA AAC TAC
 144
 Leu Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn Tyr
 35 40 45
 GTA ATG GCC AGC TTC TAC AAA GCG GAG GAG CTC TAC TAA
 183
 Val Met Ala Ser Phe Tyr Lys Ala Glu Glu Leu Tyr
 50 55 60

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 153:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 210
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

AGC CAT CTT GTC AAG TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG AAT
 48
 Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
 1 5 10 15
 GGA GGC GAG TGC TTC ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA TAC
 96
 Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
 20 25 30
 TTG TGC AAG TGC CCA AAT GAG TTT ACT GGT GAT CGC TGC CAA AAC TAC
 144
 Leu Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn Tyr
 35 40 45

- 150 -

GTA ATG GCC AGC TTC TAC AAG CAT CTT GGG ATT GAA TTT ATG GAG AAA
 192
 Val Met Ala Ser Phe Tyr Lys His Leu Gly Ile Glu Phe Met Glu Lys
 50 55 60

GCG GAG GAG CTC TAC TAA
 210
 Ala Glu Glu Leu Tyr
 65

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 154:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 267
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

AGC CAT CTT GTC AAG TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG AAT
 48
 Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
 1 5 10 15

GGA GGC GAG TGC TTC ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA TAC
 96
 Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
 20 25 30

TTG TGC AAG TGC CAA CCT GGA TTC ACT GGA GCG AGA TGT ACT GAG AAT
 144
 Leu Cys Lys Cys Gln Pro Gly Phe Thr Gly Ala Arg Cys Thr Glu Asn
 35 40 45

GTG CCC ATG AAA GTC CAA ACC CAA GAA AAG TGC CCA AAT GAG TTT ACT
 192
 Val Pro Met Lys Val Gln Thr Gln Glu Lys Cys Pro Asn Glu Phe Thr
 50 55 60

GGT GAT CGC TGC CAA AAC TAC GTA ATG GCC AGC TTC TAC AGT ACG TCC
 240
 Gly Asp Arg Cys Gln Asn Tyr Val Met Ala Ser Phe Tyr Ser Thr Ser
 65 70 75 80

ACT CCC TTT CTG TCT CTG CCT GAA TAG
 267
 Thr Pro Phe Leu Ser Leu Pro Glu
 85

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 155:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 252
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- 151 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

AGC CAT CTT GTC AAG TGT GCA GAG AAG GAG AAA ACT TTC TGT GTG AAT
 48
 Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
 1 5 10 15
 GGA GGC GAG TGC TTC ATG GTG AAA GAC CTT TCA AAT CCC TCA AGA TAC
 96
 Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
 20 25 30
 TTG TGC AAG TGC CAA CCT GGA TTC ACT GGA GCG AGA TGT ACT GAG AAT
 144
 Leu Cys Lys Cys Gln Pro Gly Phe Thr Gly Ala Arg Cys Thr Glu Asn
 35 40 45
 GTG CCC ATG AAA GTC CAA ACC CAA GAA AAG TGC CCA AAT GAG TTT ACT
 192
 Val Pro Met Lys Val Gln Thr Gln Glu Lys Cys Pro Asn Glu Phe Thr
 50 55 60
 GGT GAT CGC TGC CAA AAC TAC GTA ATG GCC AGC TTC TAC AAA GCG GAG
 240
 Gly Asp Arg Cys Gln Asn Tyr Val Met Ala Ser Phe Tyr Lys Ala Glu
 65 70 75 80
 GAG CTC TAC TAA
 252
 Glu Leu Tyr

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 156:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 128
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

CC ACA TCC ACA TCT ACA GCT GGG ACA AGC CAT CTT GTC AAG TGT GCA 47
 Thr Ser Thr Ser Thr Ala Gly Thr Ser His Leu Val Lys Cys Ala
 1 5 10 15
 GAG AAG GAG AAA ACT TTC TGT GTG AAT GGA GGC GAG TGC TTC ATG GTG 95
 Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys Phe Met Val
 20 25 30
 AAA GAC CTT TCA AAT CCC TCA AGA TAC TTG T GC 128
 Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu
 35 40

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 157:

(i) SEQUENCE CHARACTERISTICS:

- 152 -

(A) LENGTH: 141
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

A	CAT AAC CTT ATA GCT GAG CTA AGG AGA AAC AAG GCC CAC AGA TCC	46
	His Asn Leu Ile Ala Glu Leu Arg Arg Asn Lys Ala His Arg Ser	
	1 5 10 15	
AAA	TGC ATG CAG ATC CAG CTT TCC GCA ACT CAT CTT AGA GCT TCT TCC	94
	Lys Cys Met Gln Ile Gln Leu Ser Ala Thr His Leu Arg Ala Ser Ser	
	20 25 30	
ATT	CCC CAT TGG GCT TCA TTC TCT AAG ACC CCT TGG CCT TTA GGA AG	141
	Ile Pro His Trp Ala Ser Phe Ser Lys Thr Pro Trp Pro Leu Gly Arg	
	35 40 45	

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 158:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in positions 15 and 22 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

Ala	Ala Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Xaa Phe
1	5 10 15
Met	Val Lys Asp Leu Xaa Asn Pro
	20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 159:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 745
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

ATG	AGA TGG CGA CGC GCC CCG CGC CGC TCC GGG CGT CCC GGC CCC CGG
48	
Met	Arg Trp Arg Arg Ala Pro Arg Arg Ser Gly Arg Pro Gly Pro Arg
1	5 10 15

- 153 -

GCC CAG CGC CCC GGC TCC GCC GCC CGC TCG TCG CCG CCG CTG CCG CTG
 96
 Ala Gln Arg Pro Gly Ser Ala Ala Arg Ser Ser Pro Pro Leu Pro Leu
 20 25 30

CTG CCA CTA CTG CTG CTG CTG GGG ACC GCG GCC CTG GCG CCG GGG GCG
 144
 Leu Pro Leu Leu Leu Leu Leu Gly Thr Ala Ala Leu Ala Pro Gly Ala
 35 40 45

GCG GCC GGC AAC GAG GCG GCT CCC GCG GGG GCC TCG GTG TGC TAC TCG
 192
 Ala Ala Gly Asn Glu Ala Ala Pro Ala Gly Ala Ser Val Cys Tyr Ser
 50 55 60

TCC CCG CCC AGC GTG GGA TCG GTG CAG GAG CTA GCT CAG CGC GCC GCG
 240
 Ser Pro Pro Ser Val Gly Ser Val Gln Glu Leu Ala Gln Arg Ala Ala
 65 70 75 80

GTG GTG ATC GAG GGA AAG GTG CAC CCG CAG CGG CGG CAG CAG GGG GCA
 288
 Val Val Ile Glu Gly Lys Val His Pro Gln Arg Arg Gln Gln Gly Ala
 85 90 95

CTC GAC AGG AAG GCG GCG GCG GCG GCG GCG GAG GCA GGG GCG TGG GGC
 336
 Leu Asp Arg Lys Ala Ala Ala Ala Ala Gly Glu Ala Gly Ala Trp Gly
 100 105 110

GGC GAT CGC GAG CCG CCA GCC GCG GGC CCA CGG GCG CTG GGG CCG CCC
 384
 Gly Asp Arg Glu Pro Pro Ala Ala Gly Pro Arg Ala Leu Gly Pro Pro
 115 120 125

GCC GAG GAG CCG CTG CTC GCC GCC AAC GGG ACC GTG CCC TCT TGG CCC
 432
 Ala Glu Glu Pro Leu Leu Ala Ala Asn Gly Thr Val Pro Ser Trp Pro
 130 135 140

ACC GCC CCG GTG CCC AGC GCC GGC GAG CCC GGG GAG GAG GCG CCC TAT
 480
 Thr Ala Pro Val Pro Ser Ala Gly Glu Pro Gly Glu Glu Ala Pro Tyr
 145 150 155 160

CTG GTG AAG GTG CAC CAG GTG TGG GCG GTG AAA GCC GGG GGC TTG AAG
 528
 Leu Val Lys Val His Gln Val Trp Ala Val Lys Ala Gly Gly Leu Lys
 165 170 175

AAG GAC TCG CTG CTC ACC GTG CGC CTG GGG ACC TGG GGC CAC CCC GCC
 576
 Lys Asp Ser Leu Leu Thr Val Arg Leu Gly Thr Trp Gly His Pro Ala
 180 185 190

TTC CCC TCC TGC GGG AGG CTC AAG GAG GAC AGC AGG TAC ATC TTC TTC
 624
 Phe Pro Ser Cys Gly Arg Leu Lys Glu Asp Ser Arg Tyr Ile Phe Phe
 195 200 205

- 154 -

ATG GAG CCC GAC GCC AAC AGC ACC AGC CGC GCG CCG GCC GCC TTC CGA
 672
 Met Glu Pro Asp Ala Asn Ser Thr Ser Arg Ala Pro Ala Ala Phe Arg
 210 215 220
 GCC TCT TTC CCC CCT CTG GAG ACG GGC CGG AAC CTC AAG AAG GAG GTC
 720
 Ala Ser Phe Pro Pro Leu Glu Thr Gly Arg Asn Leu Lys Lys Glu Val
 225 230 235 240
 AGC CGG GTG CTG TGC AAG CGG TGC G
 745
 Ser Arg Val Leu Cys Lys Arg Cys
 245

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 160:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

Xaa Ala Leu Ala Ala Ala Gly Tyr Asp Val Glu Lys
 1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 161:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (D) OTHER INFORMATION: Xaa in position 1 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

Xaa Leu Val Leu Arg
 1 5

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 162:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11
- (B) TYPE: amino acid
- (C) STRANDEDNESS:

- 155 -

(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: Xaa in positions 1, 2, and 3 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

Xaa Xaa Xaa Tyr Pro Gly Gln Ile Thr Ser Asn
1 5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 163:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N in positions 25 and 36 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

ATAGGGAAGG GCGGGGAAG GGTCCCTC NGCAGGGCCG GGCTTGCCTC TGGAGCCTCT
60

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 164:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: N in position 16 is unknown.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

TTTACACATA TATTCNCC
18

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 165:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

- 156 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

Glu Thr Gln Pro Asp Pro Gly Gln Ile Leu Lys Lys Val Pro Met Val
 1 5 10 15
 Ile Gly Ala Tyr Thr
 20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 166:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 422
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

Met Arg Trp Arg Arg Ala Pro Arg Arg Ser Gly Arg Pro Gly Pro Arg
 1 5 10 15
 Ala Gln Arg Pro Gly Ser Ala Ala Arg Ser Ser Pro Pro Leu Pro Leu
 20 25 30
 Leu Pro Leu Leu Leu Leu Gly Thr Ala Ala Leu Ala Pro Gly Ala
 35 40 45
 Ala Ala Gly Asn Glu Ala Ala Pro Ala Gly Ala Ser Val Cys Tyr Ser
 50 55 60
 Ser Pro Pro Ser Val Gly Ser Val Gln Glu Leu Ala Gln Arg Ala Ala
 65 70 75 80
 Val Val Ile Glu Gly Lys Val His Pro Gln Arg Arg Gln Gln Gly Ala
 85 90 95
 Leu Asp Arg Lys Ala Ala Ala Ala Ala Gly Glu Ala Gly Ala Trp Gly
 100 105 110
 Gly Asp Arg Glu Pro Pro Ala Ala Gly Pro Arg Ala Leu Gly Pro Pro
 115 120 125
 Ala Glu Glu Pro Leu Leu Ala Ala Asn Gly Thr Val Pro Ser Trp Pro
 130 135 140
 Thr Ala Pro Val Pro Ser Ala Gly Glu Pro Gly Glu Glu Ala Pro Tyr
 145 150 155 160
 Leu Val Lys Val His Gln Val Trp Ala Val Lys Ala Gly Gly Leu Lys
 165 170 175
 Lys Asp Ser Leu Leu Thr Val Arg Leu Gly Thr Trp Gly His Pro Ala
 180 185 190
 Phe Pro Ser Cys Gly Arg Leu Lys Glu Asp Ser Arg Tyr Ile Phe Phe
 195 200 205
 Met Glu Pro Asp Ala Asn Ser Thr Ser Arg Ala Pro Ala Ala Phe Arg
 210 215 220

- 157 -

Ala Ser Phe Pro Pro Leu Glu Thr Gly Arg Asn Leu Lys Lys Glu Val
 225 230 235 240
 Ser Arg Val Leu Cys Lys Arg Cys Ala Leu Pro Pro Gln Leu Lys Glu
 245 250 255
 Met Lys Ser Gln Glu Ser Ala Ala Gly Ser Lys Leu Val Leu Arg Cys
 260 265 270
 Glu Thr Ser Ser Glu Tyr Ser Ser Leu Arg Phe Lys Trp Phe Lys Asn
 275 280 285
 Gly Asn Glu Leu Asn Arg Lys Asn Lys Pro Gln Asn Ile Lys Ile Gln
 290 295 300
 Lys Lys Pro Gly Lys Ser Glu Leu Arg Ile Asn Lys Ala Ser Leu Ala
 305 310 315 320
 Asp Ser Gly Glu Tyr Met Cys Lys Val Ile Ser Lys Leu Gly Asn Asp
 325 330 335
 Ser Ala Ser Ala Asn Ile Thr Ile Val Glu Ser Asn Ala Thr Ser Thr
 340 345 350
 Ser Thr Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys
 355 360 365
 Thr Phe Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser
 370 375 380
 Asn Pro Ser Arg Tyr Leu Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp
 385 390 395 400
 Arg Cys Gln Asn Tyr Val Met Ala Ser Phe Tyr Ser Thr Ser Thr Pro
 405 410 415
 Phe Leu Ser Leu Pro Glu
 420

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 167:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

Met Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys Lys
 1 5 10 15
 Glu Arg Gly Ser Gly Lys Lys Pro Glu Ser Ala Ala Gly Ser Gln Ser
 20 25 30
 Pro Arg Glu Ile Ile Thr Gly Met Pro Ala Ser Thr Glu Gly Ala Tyr
 35 40 45

- 158 -

Val Ser Ser Glu Ser Pro Ile Arg Ile Ser Val Ser Thr Glu Gly Ala
50 55 60

Asn Thr Ser Ser Ser
65

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 168:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

Arg Lys Gly Asp Val Pro Gly Pro Arg Val Lys Ser Ser Arg Ser Thr
1 5 10 15

Thr Thr Ala

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 169:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 231
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

CGCGAGCGCC TCAGCGGGC CGCTCGCTCT CCCCCTCGAG GGACAAACTT TTCCCAAACC
60

CGATCCGAGC CCTTGGACCA AACTCGCCTG CGCCGAGAGC CGTCCGCGTA GAGCGCTCCG
120

TCTCCGGCGA GATGTCCGAG CGCAAAGAAG GCAGAGGCCA AGGGAAGGGC AAGAAGAAGG
180

AGCGAGGCTC CGGCAAGAAG CCGGAGTCCG CGGCGGGCAG CCAGAGCCCA G
231

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 170:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 178
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

CCTTGCCCTCC CCGATTGAAA GAGATGAAA GCCAGGAATC GGCTGCAGGT TCCAAACTAG
60

- 159 -

TCCTTCGGTG TGAAACCAGT TCTGAATACT CCTCTCTCAG ATTCAAGTGG TTCAAGAATG
120

GGAATGAATT GAATCGAAAA AACAAACCAC AAAATATCAA GATACAAAAA AAGCCAGG
178

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 171:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 122
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

GAAGTCAGAA CTTGCGATTA ACAAAGCATC ACTGGCTGAT TCTGGAGAGT ATATGTGCAA
60

AGTGATCAGC AAATTAGGAA ATGACAGTGC CTCTGCCAAT ATCACCATCG TGGAATCAAA
120

CG
122

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 172:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 102
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

AGATCATCAC TGGTATGCCA GCCTCAACTG AAGGAGCATA TGTGTCTTCA GAGTCTCCCA
60

TTAGAATATC AGTATCCACA GAAGGAGCAA ATACTTCTTC AT
102

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 173:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 128
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

CTACATCTAC ATCCACCACT GGGACAAGCC ATCTTGTAAG ATGTGCGGAG AAGGAGAAAA
60

- 160 -

CTTCTGTGT GAATGGAGGG GAGTGCTTCA TGGTGAAAGA CCTTCAAAC CCCTCGAGAT
120

ACTTGTGC
128

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 174:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

AAGTGCCAAC CTGGATTCAC TGGAGCAAGA TGTACTGAGA ATGTGCCCAT GAAAGTCCAA
60

AACCAAGAA
69

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 175:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 60
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

AAGTGCCCAA ATGAGTTTAC TGGTGATCGC TGCCAAACT ACGTAATGGC CAGCTTCTAC
60

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 176:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

AGTACGTCCA CTCCCTTCT GTCTCTGCCT GAATAG
36

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 177:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 569
- (B) TYPE: nucleic acid

- 161 -

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

AAGGCGGAGG AGCTGTACCA GAAGAGAGTG CTGACCATAA CCGGCATCTG CATCGCCCTC
60
CTTGTGGTCG GCATCATGTG TGTGGTGGCC TACTGCAAAA CCAAGAAACA GCGGAAAAAG
120
CTGCATGACC GTCTTCGGCA GAGCCTTCGG TCTGAACGAA ACAATATGAT GAACATTGCC
180
AATGGGCCTC ACCATCCTAA CCCACCCCCC GAGAATGTCC AGCTGGTGAA TCAATACGTA
240
TCTAAAAACG TCATCTCCAG TGAGCATATT GTTGAGAGAG AAGCAGAGAC ATCCTTTTCC
300
ACCACTCACT ATACTTCAC AGCCCATCAC TCCACTACTG TCACCCAGAC TCCTAGCCAC
360
AGCTGGAGCA ACGGACACAC TGAAAGCATC CTTTCCGAAA GCCACTCTGT AATCGTGATG
420
TCATCCGTAG AAAACAGTAG GCACAGCAGC CCAACTGGGG GCCCAAGAGG ACGTCTTAAT
480
GGCACAGGAG GCCCTCGTGA ATGTAACAGC TTCCTCAGGC ATGCCAGAGA AACCCCTGAT
540
TCCTACCGAG ACTCTCCTCA TAGTGAAAG
569

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 178:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 730
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

GTATGTGTCA GCCATGACCA CCCCGGCTCG TATGTCACCT GTAGATTTC ACACGCCAAG
60
CTCCCCCAA TCGCCCCCTT CGGAAATGTC TCCACCCGTG TCCAGCATGA CGGTGTCCAT
120
GCCTTCCATG GCGGTCAGCC CCTTCATGGA AGAAGAGAGA CCTCTACTTC TCGTGACACC
180
ACCAAGGCTG CGGGAGAAGA AGTTTGACCA TCACCCTCAG CAGTTCAGCT CCTTCCACCA
240
CAACCCGCG CATGACAGTA ACAGCCTCCC TGCTAGCCCC TTGAGGATAG TGGAGGATGA
300

- 162 -

GGAGTATGAA ACGACCCAAG AGTACGAGCC AGCCCAAGAG CCTGTTAAGA AACTCGCCAA
360

TAGCCGGCGG GCCAAAAGAA CCAAGCCCAA TGGCCACATT GCTAACAGAT TGGAAGTGGG
420

CAGCAACACA AGCTCCCAGA GCAGTAATC AGAGAGTGAA ACAGAAGATG AAAGAGTAGG
480

TGAAGATACG CCTTTCCTGG GCATACAGAA CCCCTGGCA GCCAGTCTTG AGGCAACACC
540

TGCCTTCCGC CTGGCTGACA GCAGGACTAA CCCAGCAGGC CGCTTCTCGA CACAGGAAGA
600

AATCCAGGCC AGGCTGTCTA GTGTAATTGC TAACCAAGAC CCTATTGCTG TATAAACCT
660

AAATAAACAC ATAGATTCAC CTGTAAACT TTATTTTATA TAATAAAGTA TTCCACCTTA
720

AATTAAACAA
730

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 179:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

TCGGGCTCCA TGAAGAAGAT GTA
23

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 180:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

TCCATGAAGA AGATGTACCT GCT
23

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 181:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22
(B) TYPE: nucleic acid

- 163 -

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

ATGTACCTGC TGCCTCCTT GA
22

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 182:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

TTGAAGAAGG ACTCGCTGCT CA
22

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 183:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

AAAGCCGGGG GCTTGAAGAA
20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 184:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

ATGARGTGTG GGCGGCCGAAA
20

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 185:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 422
(B) TYPE: amino acid
(C) STRANDEDNESS:

- 164 -

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

Met Arg Trp Arg Arg Ala Pro Arg Arg Ser Gly Arg Pro Gly Pro Arg
 1 5 10 15
 Ala Gln Arg Pro Gly Ser Ala Ala Arg Ser Ser Pro Pro Leu Pro Leu
 20 25 30
 Leu Pro Leu Leu Leu Leu Gly Thr Ala Ala Leu Ala Pro Gly Ala
 35 40 45
 Ala Ala Gly Asn Glu Ala Ala Pro Ala Gly Ala Ser Val Cys Tyr Ser
 50 55 60
 Ser Pro Pro Ser Val Gly Ser Val Gln Glu Leu Ala Gln Arg Ala Ala
 65 70 75 80
 Val Val Ile Glu Gly Lys Val His Pro Gln Arg Arg Gln Gln Gly Ala
 85 90 95
 Leu Asp Arg Lys Ala Ala Ala Ala Gly Glu Ala Gly Ala Trp Gly
 100 105 110
 Gly Asp Arg Glu Pro Pro Ala Ala Gly Pro Arg Ala Leu Gly Pro Pro
 115 120 125
 Ala Glu Glu Pro Leu Leu Ala Ala Asn Gly Thr Val Pro Ser Trp Pro
 130 135 140
 Thr Ala Pro Val Pro Ser Ala Gly Glu Pro Gly Glu Glu Ala Pro Tyr
 145 150 155 160
 Leu Val Lys Val His Gln Val Trp Ala Val Lys Ala Gly Gly Leu Lys
 165 170 175
 Lys Asp Ser Leu Leu Thr Val Arg Leu Gly Thr Trp Gly His Pro Ala
 180 185 190
 Phe Pro Ser Cys Gly Arg Leu Lys Glu Asp Ser Arg Tyr Ile Phe Phe
 195 200 205
 Met Glu Pro Asp Ala Asn Ser Thr Ser Arg Ala Pro Ala Ala Phe Arg
 210 215 220
 Ala Ser Phe Pro Pro Leu Glu Thr Gly Arg Asn Leu Lys Lys Glu Val
 225 230 235 240
 Ser Arg Val Leu Cys Lys Arg Cys Ala Leu Pro Pro Gln Leu Lys Glu
 245 250 255
 Met Lys Ser Gln Glu Ser Ala Ala Gly Ser Lys Leu Val Leu Arg Cys
 260 265 270
 Glu Thr Ser Ser Glu Tyr Ser Ser Leu Arg Phe Lys Trp Phe Lys Asn
 275 280 285
 Gly Asn Glu Leu Asn Arg Lys Asn Lys Pro Gln Asn Ile Lys Ile Gln
 290 295 300

- 165 -

Lys Lys Pro Gly Lys Ser Glu Leu Arg Ile Asn Lys Ala Ser Leu Ala
 305 310 315 320
 Asp Ser Gly Glu Tyr Met Cys Lys Val Ile Ser Lys Leu Gly Asn Asp
 325 330 335
 Ser Ala Ser Ala Asn Ile Thr Ile Val Glu Ser Asn Ala Thr Ser Thr
 340 345 350
 Ser Thr Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys
 355 360 365
 Thr Phe Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser
 370 375 380
 Asn Pro Ser Arg Tyr Leu Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp
 385 390 395 400
 Arg Cys Gln Asn Tyr Val Met Ala Ser Phe Tyr Ser Thr Ser Thr Pro
 405 410 415
 Phe Leu Ser Leu Pro Glu
 420

What is claimed is:

- 166 -

1. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a polypeptide encoded by pGGF2HBS5 deposited with the A.T.C.C. November 6, 1992 (A.T.C.C. Deposit No. 75347) with a pharmaceutical carrier.

2. A method of making a medicament for the treating of muscle cells of a mammal, said method comprising admixing a polypeptide encoded by the E sequence (SEQ ID Nos. 133 and 159) and at least a portion of the peptide encoded by the DNA sequences flanking the E encoding sequence on clone pGGF2HBS5, deposited with the A.T.C.C. November 6, 1992 (A.T.C.C. Deposit No. 75347).

3. A method of making a medicament for the treating of muscle cells of a mammal, said method comprising admixing a polypeptide defined by the formula

YBAZCX

wherein YBAZCX is composed of the polypeptide segments shown in Fig. 30 (SEQ ID Nos. 133-135, 156, 159); wherein Y comprises polypeptide segment E, or is absent; wherein Z comprises polypeptide segment G or is absent; and wherein X comprises polypeptide segments C/D HKL, C/D H, C/D HL, C/D D, C/D' HL, C/D' HKL, C/D' H, C/D' D, C/D C/D' HKL, C/D C/D' H, C/D C/D' HL, C/D C/D' D, C/D D' H, C/D D' HL, C/D D' HKL, C/D' D' H, C/D' D' HL, C/D' D' HKL, C/D C/D' D' H, C/D C/D' D' HL, or C/D C/D' D' HKL with a pharmaceutical carrier.

- 167 -

4. A method of making a medicament for the treating of muscle cells of a mammal, said method comprising admixing a polypeptide defined by the formula

WBAZCX

5 wherein WBAZCX is composed of the polypeptide segments shown in Fig. 30 (SEQ ID Nos. 132, 134, 135, 137-139, 156); wherein W comprises polypeptide segment F, or is absent; wherein Z comprises polypeptide segment G or is absent; and wherein X comprises polypeptide
 10 segments C/D HKL, C/D H, C/D HL, C/D D, C/D' HL, C/D' HKL, C/D' H, C/D' D, C/D C/D' HKL, C/D C/D' H, C/D C/D' HL, C/D C/D' D, C/D D' H, C/D D' HL, C/D D' HKL, C/D' D' H, C/D' D' HL, C/D' D' HKL, C/D C/D' D' H, C/D C/D' D' HL, or C/D C/D' D' HKL with a pharmaceutical carrier.

15 5. The method of any one of claims 1-3, wherein 50 N-terminal amino acids are cleaved from said peptide comprising the E sequence (SEQ ID Nos. 133 and 159).

6. The method of claim 3 or 4, wherein X is C/D HKL.

20 7. The method of claim 3 or 4, wherein X is C/D H.

8. The method of claim 3 or 4, wherein X is C/D HL.

25 9. The method of claim 3 or 4, wherein X is C/D D.

10. The method of claim 3 or 4, wherein X is C/D' HL.

- 168 -

11. The method of claim 3 or 4, wherein X is C/D'
HKL.
12. The method of claim 3 or 4, wherein X is C/D'
H.
- 5 13. The method of claim 3 or 4, wherein X is C/D'
D.
14. The method of claim 3 or 4, wherein X is C/D
C/D' HKL.
- 10 15. The method of claim 3 or 4, wherein X is C/D
C/D' H.
16. The method of claim 3 or 4, wherein X is C/D
C/D' HL.
17. The method of claim 3 or 4, wherein X is C/D
C/D' D.
- 15 18. The method of claim 3 or 4, wherein X is C/D
D' H.
19. The method of claim 3 or 4, wherein X is C/D
D' HL.
- 20 20. The method of claim 3 or 4, wherein X is C/D
D' HKL.
21. The method of claim 3 or 4, wherein X is C/D'
D' H.

- 169 -

22. The method of claim 3 or 4, wherein X is C/D'
D' HL.

23. The method of claim 3 or 4, wherein X is C/D'
D' HKL.

5 24. The method of claim 3 or 4, wherein X is C/D
C/D' D' H.

25. The method of claim 3 or 4, wherein X is C/D
C/D' D' HL.

10 26. The method of claim 3 or 4, wherein X is C/D
C/D' D' HKL.

27. A method of making a medicament for the
treating of muscle cells of a mammal said method
comprising admixing a polypeptide comprising FBA
polypeptide segments having the amino acid sequences
15 shown in Fig. 30 (SEQ ID Nos. 132, 134, 135) with a
pharmaceutically acceptable carrier.

28. A method of making a medicament for the
treating of muscle cells of a mammal said method
comprising admixing a polypeptide comprising FBA'
20 polypeptide segments having the amino acid sequences
shown in Fig. 30 (SEQ ID Nos. 132, 134, 136) with a
pharmaceutically acceptable carrier.

29. A method of making a medicament for the
treating of muscle cells of a mammal said method
25 comprising admixing a polypeptide comprising FEBA
polypeptide segments having the amino acid sequences
shown in Fig. 30 (SEQ ID Nos. 132, 135, 159) with a
pharmaceutically acceptable carrier.

- 170 -

30. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a polypeptide comprising FEBA' polypeptide segments having the amino acid sequences
5 corresponding to polypeptide segments shown in Fig. 30 (SEQ ID Nos. 132-134, 136, 159) to muscle cells with a pharmaceutically acceptable carrier.

31. A method of making a medicament for the treating of muscle cells of a mammal said method
10 comprising admixing GGF2 polypeptide with a pharmaceutically acceptable carrier.

32. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a compound which specifically binds
15 the p185^{erbB2} receptor of muscle cells with a pharmaceutically acceptable carrier.

33. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a polypeptide comprising EGFL1,
20 having the amino acid sequence shown Fig. 37, Seq. ID No. 150, with a pharmaceutically acceptable carrier.

34. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a polypeptide comprising EGFL2,
25 having the amino acid sequence shown in Fig. 38, Seq. ID No. 151, with a pharmaceutically acceptable carrier.

- 171 -

35. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a polypeptide comprising EGFL3, with the amino acid sequence shown in Fig. 39, Seq. ID No. 152, with a pharmaceutically acceptable carrier.

36. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a polypeptide comprising EGFL4, with the amino acid sequence shown in Fig. 40, Seq. ID No. 153, with a pharmaceutically acceptable carrier.

37. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a polypeptide comprising EGFL5, with the amino acid sequence shown in Fig. 41, Seq. ID No. 154, to muscle cells, with a pharmaceutically acceptable carrier.

38. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a polypeptide, comprising EGFL6, with the amino acid sequence shown Fig. 42, Seq. ID No. 155, with a pharmaceutically acceptable carrier.

39. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a 35 kD polypeptide factor isolated from the rat I-EJ transformed fibroblast cell line to said muscle cells, with a pharmaceutically acceptable carrier.

- 172 -

40. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a 75 kD polypeptide factor isolated from the SKBR-3 human breast cell line to said muscle
5 cells, with a pharmaceutically acceptable carrier.

41. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a 44 kD polypeptide factor isolated from the rat I-EJ transformed fibroblast cell line to
10 said muscle cells, with a pharmaceutically acceptable carrier.

42. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a 45 kD polypeptide factor isolated
15 from the MDA - MB 231 human breast cell line to said muscle cells, with a pharmaceutically acceptable carrier.

43. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a 7 to 14 kD polypeptide factor
20 isolated from the ATL-2 human T-cell line to said muscle cells, with a pharmaceutically acceptable carrier.

44. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a 25 kD polypeptide factor isolated
25 from activated mouse peritoneal macrophages to said muscle cells, with a pharmaceutically acceptable carrier.

- 173 -

45. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a 25 kD polypeptide factor isolated from bovine kidney to said muscle cells, with a
5 pharmaceutically acceptable carrier.

46. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a ARIA polypeptide to said muscle cells, with a pharmaceutically acceptable carrier.

10 47. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing a 46-47 kD polypeptide factor which stimulates 0-2A glial progenitor cells to said muscle cells, with a pharmaceutically acceptable carrier.

15 48. A method of making a medicament for the treating of muscle cells of a mammal said method comprising admixing GGF-III to said muscle cells, with a pharmaceutically acceptable carrier.

- 174 -

49. A method of making a medicament for the treating of muscle cells of a mammal, said method comprising admixing with a pharmaceutically acceptable carrier, a DNA sequence encoding a polypeptide of the
5 formula

YBAZCX

wherein YBAZCX is composed of the polypeptide segments shown in Fig. 30 (SEQ ID Nos. 133-135, 156, 159); wherein Y comprises polypeptide segment E, or is
10 absent; wherein Z comprises polypeptide segment G or is absent; and wherein X comprises polypeptide segments C/D HKL, C/D H, C/D HL, C/D D, C/D' HL, C/D' HKL, C/D' H, C/D' D, C/D C/D' HKL, C/D C/D' H, C/D C/D' HL, C/D C/D' D, C/D D' H, C/D D' HL, C/D D' HKL, C/D' D' H, C/D' D'
15 HL, C/D' D' HKL, C/D C/D' D' H, C/D C/D' D' HL, or C/D C/D' D' HKL, said DNA in an expressible genetic construction.

- 175 -

50. A method of making a medicament for the treating of muscle cells of a mammal, said method comprising admixing with a pharmaceutically acceptable carrier, a DNA sequence encoding a polypeptide of the formula

WBAZCX

wherein WBAZCX is composed of the polypeptide segments shown in Fig. 30 (SEQ ID Nos. 132, 134, 135, 137-139, 156); wherein W comprises polypeptide segment F, or is absent; wherein Z comprises polypeptide segment G or is absent; and wherein X comprises polypeptide segments C/D HKL, C/D H, C/D HL, C/D D, C/D' HL, C/D' HKL, C/D' H, C/D' D, C/D C/D' HKL, C/D C/D' H, C/D C/D' HL, C/D C/D' D, C/D D' H, C/D D' HL, C/D D' HKL, C/D' D' H, C/D' D' HL, C/D' D' HKL, C/D C/D' D' H, C/D C/D' D' HL, or C/D C/D' D' HKL, said DNA in an expressible genetic construction.

51. A method of making a medicament for the prophylaxis or treatment of pathophysiological condition of the musculature in a mammal in which said condition involves a muscle cell type which is sensitive or responsive to a polypeptide as defined in any one of claims 1, 3, 4, and 31, said method comprising admixing an effective amount of said polypeptide with a pharmaceutically acceptable carrier.

52. A method of making a medicament for the treatment of a condition which involves muscle damage in a mammal, said method comprising admixing an effective amount of a polypeptide, as defined in any one of claims 1, 3, 4, and 31 with a pharmaceutically acceptable carrier.

- 176 -

53. The method of any one of claims 1, 3, 4, and 31, wherein said medicament is for decreasing the atrophy of said muscle cells.

54. The method of any one of claims 1, 3, 4, and 31, wherein said medicament is for increasing the muscle fibers present in said mammal.

55. The method of any one of claims 1, 3, 4, and 31, wherein said medicament is for increasing muscle cell survival in a said mammal.

56. The method of any one of claims 1, 3, 4, and 31, wherein said medicament is for increasing muscle growth in a said mammal.

57. The method of any one of claims 1, 3, 4, and 31, wherein said medicament is for increasing muscle regeneration in a said mammal.

58. The method of any one of claims 1, 3, 4, and 31, wherein said medicament is for stimulating muscle cell mitogenesis.

59. The method of any one of claims 1, 3, 4, and 31, wherein said medicament is for increasing acetylcholine receptor synthesis.

60. The method of any one of claims 1, 3, 4, and 31, wherein said medicament is for treating a patient lacking a neurotrophic factor.

- 177 -

61. A method of claims 1, 3, 4, and 31, wherein said medicament is for treating a muscle cell which is a myoblast.

62. A method of claims 1, 3, 4, and 31, wherein
5 said medicament is for treating a muscle cell which is a satellite cell.

63. A method of claims 1, 3, 4, and 31 wherein said medicament is for treating a muscle cell in skeletal muscle.

10 64. A method of claims 1, 3, 4, and 31 wherein said medicament is for treating a muscle cell in cardiac muscle.

65. A method of claims 1, 3, 4, and 31 wherein said medicament is for treating a muscle cell in smooth
15 muscle.

66. A method of claims 1, 3, 4, and 31, wherein said medicament is for treating a muscle cell in a patient with a skeletal muscle disease.

67. A method of claim 66, wherein said skeletal
20 muscle disease is a myopathy.

68. A method of claim 66, wherein said skeletal muscle disease is a dystrophy.

69. A method of claim 68, wherein said dystrophy is Duchennes muscular dystrophy.

- 178 -

70. A method of claim 68, wherein said dystrophy is Beckker's dystrophy.

71. A method of claim 66, wherein said skeletal muscle disease is a result of a neural condition.

5 72. A method of claim 66, wherein said skeletal muscle disease is an injury.

73. A method of claim 66, wherein said skeletal muscle disease is resulting from a nerve injury.

10 74. A method of claim 66, wherein said skeletal muscle disease is resulting from a neuropathy.

75. A method of claims 1, 3, 4, and 31, wherein said medicament is for treating a muscle cell in a patient with a cardiac muscle disorder.

15 76. A method of claim 75, wherein said cardiac disorder is cardiomyopathy.

77. A method of claim 75, wherein said cardiac disorder is ischemic damage.

78. A method of claim 75, wherein said cardiac disorder is a congenital disease.

20 79. A method of claim 75, wherein said cardiac disorder is cardiac trauma.

80. A method of claims 1, 3, 4, and 31, wherein said medicament is for treating a muscle cell in a patient with a smooth muscle disorder.

- 179 -

81. A method of claim 80, wherein said disorder is arterial sclerosis.

82. A method of claim 80, wherein said disorder is a vascular lesion.

5 83. A method of claim 80, wherein said disorder is a congenital vascular disease.

84. A method of claims 1, 3, 4, and 31, wherein said medicament is for treating a muscle cell which has insufficient functional acetylcholine receptors.

10 85. A method of claim 85 wherein said muscle cell lacking sufficient acetylcholine receptor is a muscle cell in a patient with myasthenia gravis.

86. A method as claimed in claim 84, wherein said condition involves muscular damage.

15 87. A method of making a medicament for the prophylaxis or treatment of a muscular tumor in a patient, said method comprising admixing an effective amount of a substance which inhibits the binding of a factor as defined in any one of claims 1, 3, 4, and 31 to
20 a receptor therefor with a pharmaceutically acceptable carrier.

88. A method of making a medicament for treating a mammal suffering from a disease of muscle cell proliferation, said method comprising admixing an
25 antibody which binds to a polypeptide of any of one of claims 1, 3, 4, and 31 with a pharmaceutically acceptable carrier.

- 180 -

89. A method of identifying a nucleic acid sequence coding for a molecule having muscle cell mitogenic activity, said method comprising contacting a cell containing sample with a muscle cell mitogen
5 specific antibody to determine expression of said mitogen in said sample and isolating said nucleic acid sequence from the cells exhibiting said expression.

90. The method of claim 31, wherein said GGF2 is human recombinant GGF2.

10 91. A method of stimulating myogenesis of a muscle cell said method comprising contacting said muscle cell with a compound which specifically binds the p185^{erbB2} receptor of muscle cells.

1/57

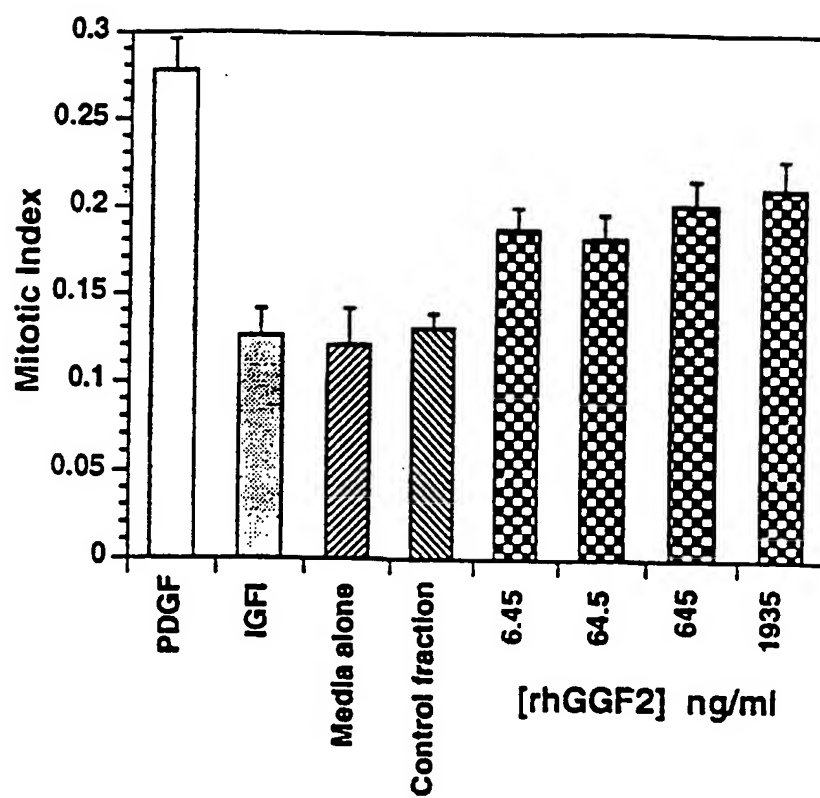


FIGURE 1

2/57

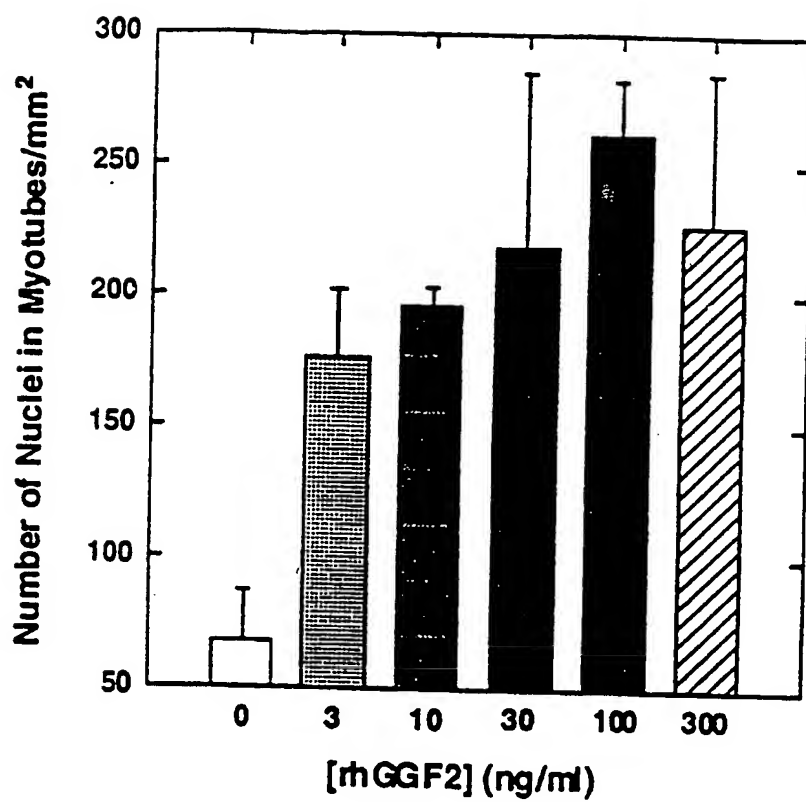


FIGURE 2

3/57

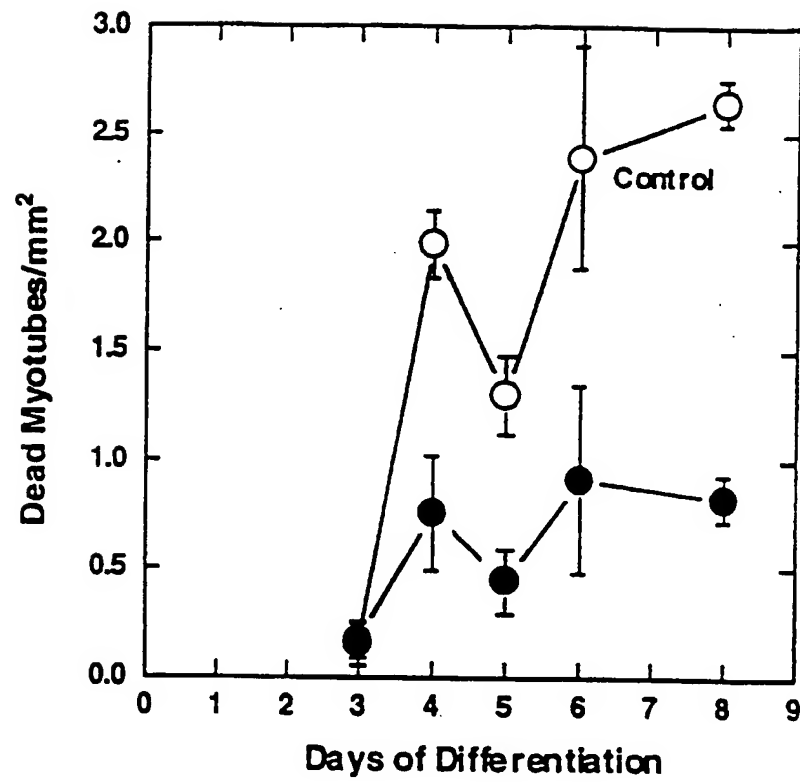


FIGURE 3

4/57

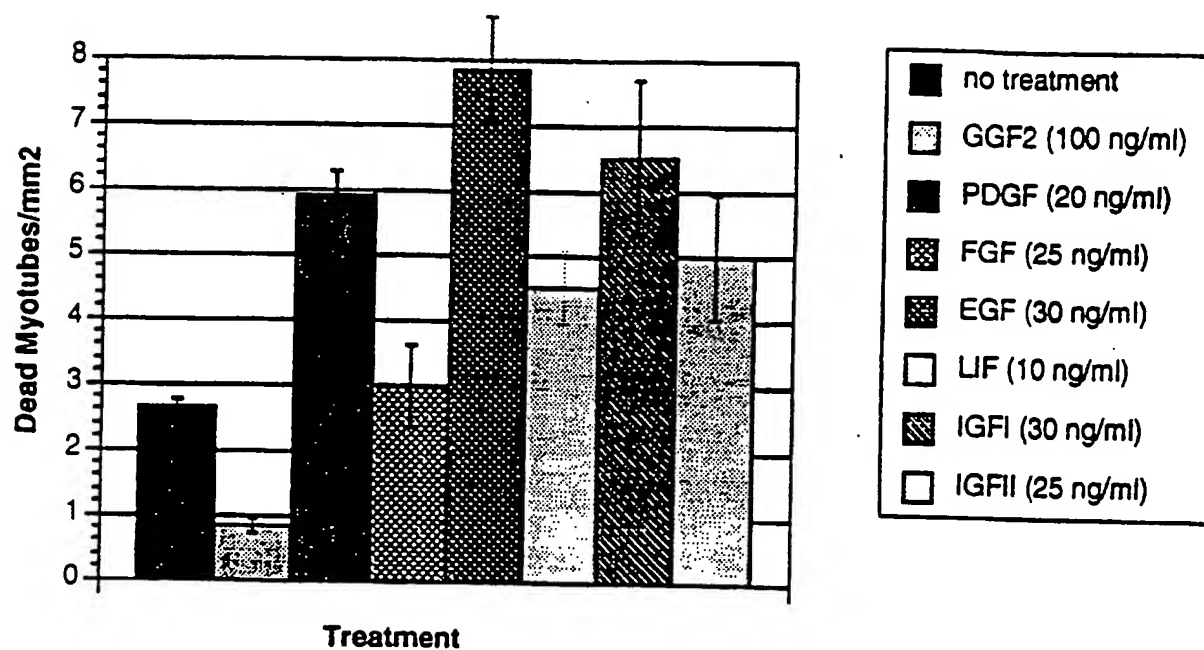


FIGURE 4

5/57

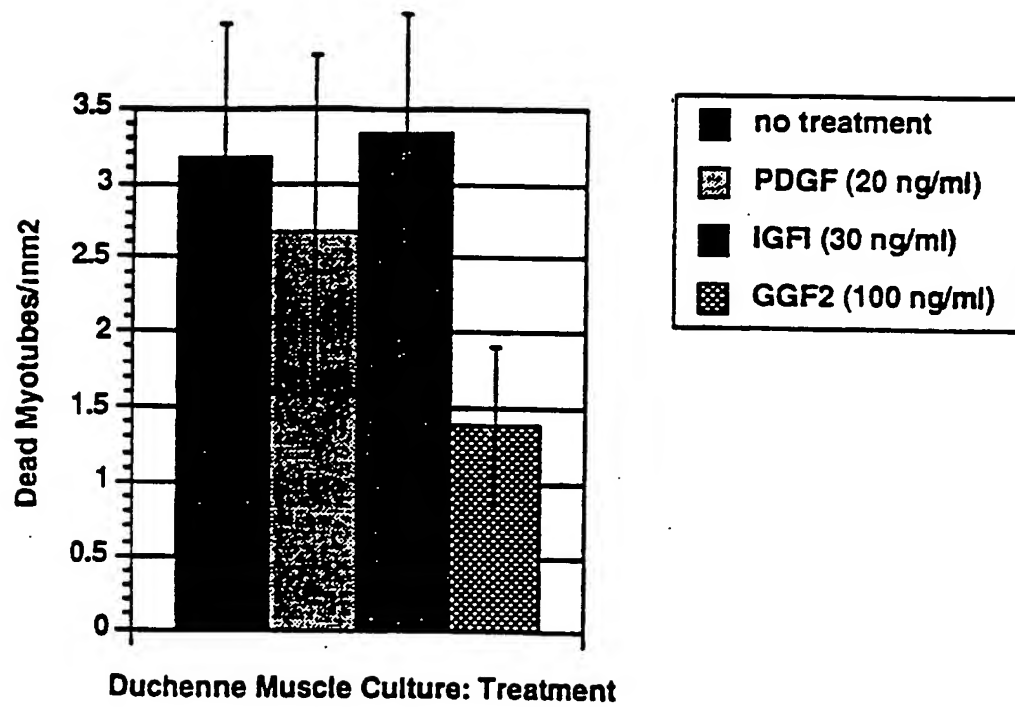


FIGURE 5

6/57

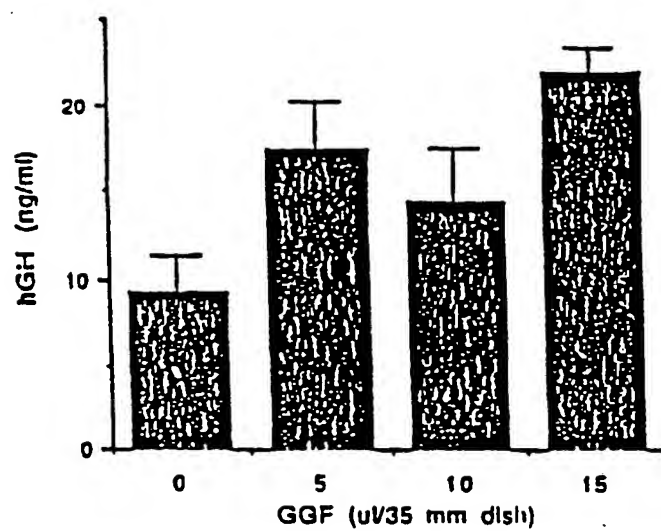
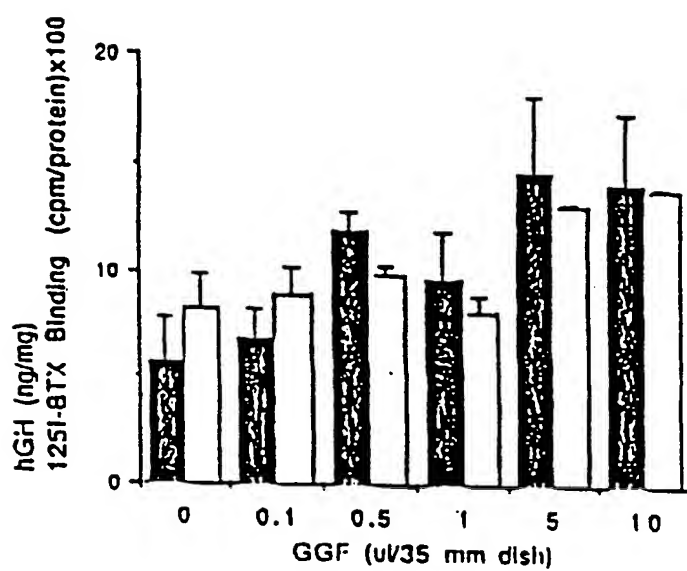


FIGURE 6

7/57

**FIGURE 7**

8/57

FIGURE 8

2F-1 01	N-terminal	F K G D A N T E (seq ID NO: 1)	
2F-1 02	Typical peptides	K R A S L A D E Y E Y M X K • (seq ID NO: 2)	
2F-1 03		K R T E T S S G L X L K • (seq ID NO: 3)	
2F-1 04		K R K L G E M W A E (seq ID NO: 4)	
2F-1 05		K R L G E K R A (seq ID NO: 5)	
2F-1 06		K R I K S E H A G L S I G D T A K • (seq ID NO: 6)	18AG-1
2F-1 07		K R A S L A D E Y E Y M R K • (seq ID NO: 7)	18AG-17
2F-1 08		K R I K G E H P G L S I G D V A K • (seq ID NO: 8)	18AG-2
2F-1 09		K R M S E Y A F F V Q T X R • (seq ID NO: 9)	18AG-1
2F-1 10		K R S E H P G L S I G D T A K • (seq ID NO: 10)	18AG-2
2F-1 11		K R A G Y F A E X A R • (seq ID NO: 11)	18AG-1
2F-1 12		K R K L E F L X A K • (seq ID NO: 12)	
2F-1 13		K R Y T E M A S E Q G A (seq ID NO: 13)	
2F-1 14		K R A K E A L A L K • (seq ID NO: 14)	
2F-1 15		K R F V L Q A K • (seq ID NO: 15)	
2F-1 16		K R L G E M W (seq ID NO: 16)	18AG-1
2F-1 17	Protease V8 peptides	E T Q P D P G Q I L K K V P M V I G A Y T (seq ID NO: 165)	
2F-1 18		E Y K C L K F K W F K K A T V M (seq ID NO: 17)	
2F-1 19		E A K Y F S K X D A (seq ID NO: 18)	
2F-1 20		E X K F Y V P (seq ID NO: 19)	Ut-alpha
2F-1 21		E L S F A S V R L P G C P P G V D P M V S F P V A L	Ut-beta (seq ID NO: 71)

9/57

FIGURE 9

A

OCF-1 01	F K G D A N T E	(SEQ ID NO: 1)
OCF-1 02	A S L A D E Y E Y M X K	(SEQ ID NO: 22)
OCF-1 03	T E T S S G L X L K	(SEQ ID NO: 23)
OCF-1 07	A S L A D E Y E Y M X K	(SEQ ID NO: 24)
OCF-1 11	A G Y F A E X A R	(SEQ ID NO: 25)
OCF-1 13	T T E M A S E Q Q A	(SEQ ID NO: 26)
OCF-1 14	A K E A L A A L K	(SEQ ID NO: 27)
OCF-1 15	F V L O A K K	(SEQ ID NO: 28)
OCF-1 17	E T O P D P G Q I L K K V P M V I G A Y T	(SEQ ID NO: 29)
OCF-1 18	E Y K C L K F K W F K K A T V M	(SEQ ID NO: 17)

B

OCF-1 20	E K K F Y V P	(SEQ ID NO: 19)
OCF-1 12	K L E F L X A K	(SEQ ID NO: 32)

10/57

FIGURE 10

	Trypsin peptides	
GGF-II 01	K/R VHQVWAAK*	(SEQ ID NO: 42)
GGF-II 02	K/R YIFFMEPEAXSSG	(SEQ ID NO: 43)
GGF-II 03	K/R LGAWGPPAFPVXY	(SEQ ID NO: 44)
GGF-II 04	K/R WfvVIEGK*	(SEQ ID NO: 45)
GGF-II 05	K/R ALAAAGYDVEK*	Histone H1 (SEQ ID NO: 160)
GGF-II 06	K/R LVLK*	(SEQ ID NO: 161)
GGF-II 07	K/R XXYPGQITSN	Trypsin (SEQ ID NO: 162)
GGF-II 08	K/R ASPVSVGSVQELVQR*	(SEQ ID NO: 46)
GGF-II 09	K/R VCLLTVAALPPT	(SEQ ID NO: 47)
GGF-II 10	K/R DLLLXY	(SEQ ID NO: 50)
	Lysyl Endopeptidase-C peptides	
GF-II 11	KVHQVWAAK*	(SEQ ID NO: 48)
GF-II 12	KASLADSGEYMXK*	(SEQ ID NO: 49)

11/57

FIGURE 11

A

GGF-II 01	VHQVWAAK	(SEQ ID NO: 42)
GGF-II 02	YIFFMEPEAXSSG	(SEQ ID NO: 43)
GGF-II 03	LGAWGPPAFPVXY	(SEQ ID NO: 44)
GGF-II 04	WFVVIEGK	(SEQ ID NO: 45)
GGF-II 08	ASPYSVGSVQELVQR	(SEQ ID NO: 46)
GGF-II 09	VCLLTVAALPPT	(SEQ ID NO: 47)
GGF-II 11	KVHQVWAAK	(SEQ ID NO: 48)
GGF-II 12	KASLADSGEYMXK	(SEQ ID NO: 49)

B

Novel Factor II Peptides - others

GGF-II 10	DLLLY	(SEQ ID NO: 5)
-----------	-------	----------------

12/57

Comparison: BrUdR JSA and [125 I]UdR x-ray method for the
DNA synthesis assay in Schwann cell cultures

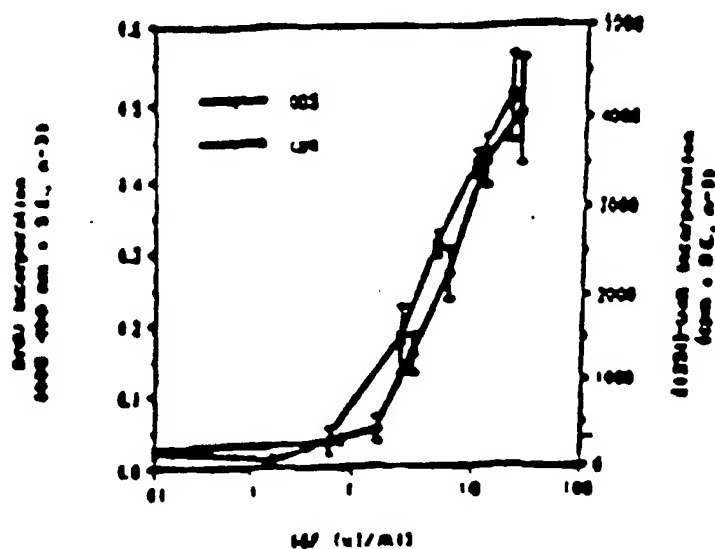


FIGURE 12

Comparison of Br-UdR immunoreactivity and
Br-UdR labelled cell number

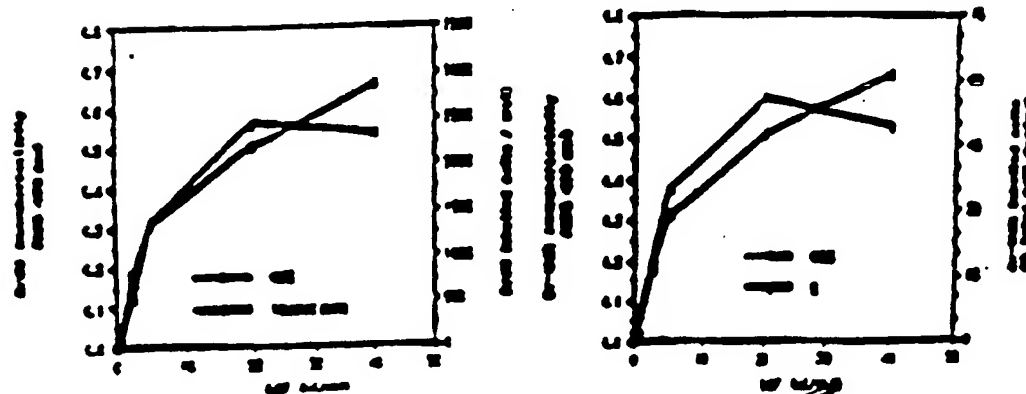


FIGURE 13A

FIGURE 13B

13/57

Mitogenic response of rat sciatic nerve Schwann cell to GGFs

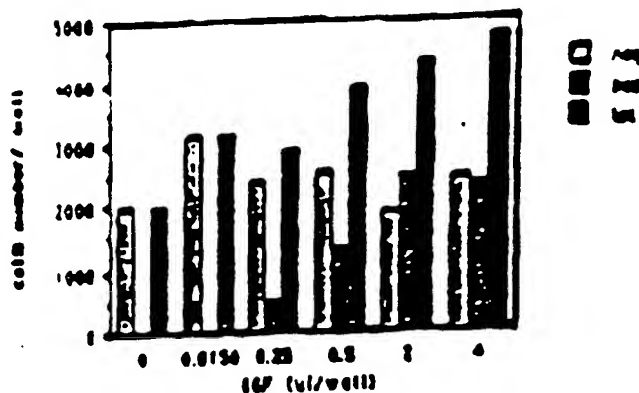


FIGURE 14

DNA synthesis in rat sciatic nerve Schwann cells and 3T3 fibroblasts in the presence of GGFs

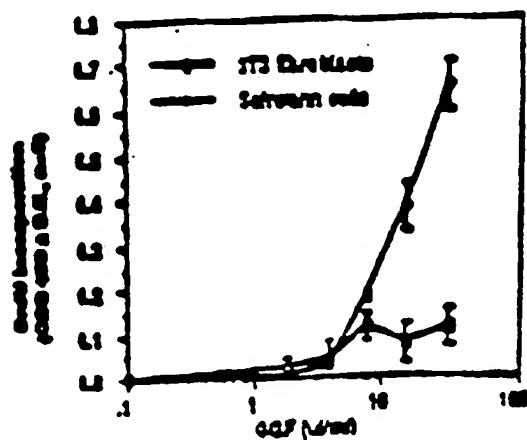


FIGURE 15

14/57

Mitogenic response of BHK₂₁ C13 cells to FCS and GGFs

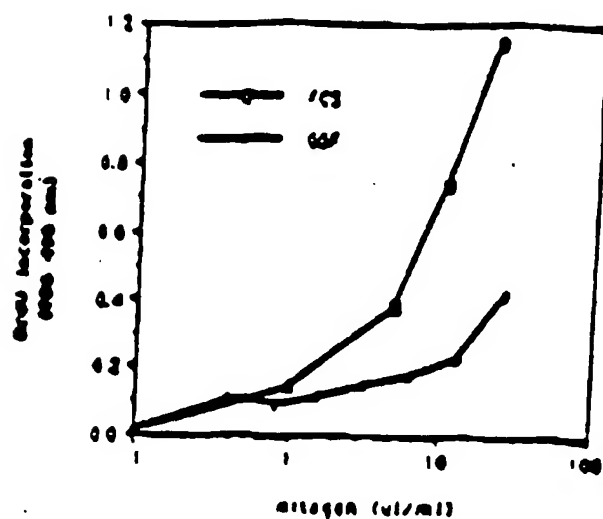


FIGURE 16

Survival and proliferation of BHK₂₁ C13 cell microcultures after 48 hours in presence of GGFs

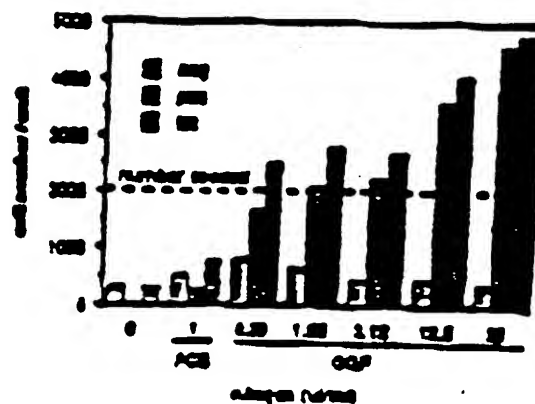


FIGURE 17

15/57

Mitogenic response of C8 cells to FCS

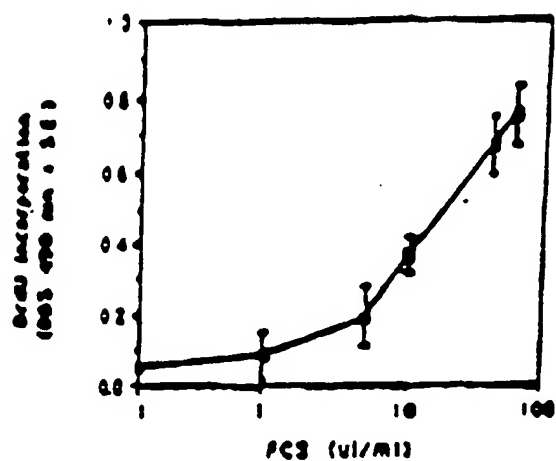


FIGURE 18

Mitogenic response of C6 cells to aFGF and GGFs

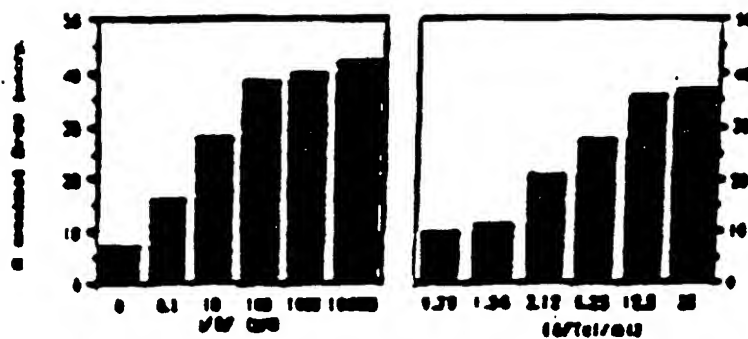


FIGURE 19

16/57

DEGENERATE OLIGONUCLEOTIDE PROBES FOR FACTOR I AND FACTOR II

Oligo	Sequence	Peptide	
535	TTYAARGGNGAYGCHAYAC!	GGFI-1	(SEQ ID NO: 51)
536	CATRTAYTCRTAYTCRTGCG!	GGFI-2	(SEQ ID NO: 52)
537	TGYTCGANGCCATYTCGFI	GGFI-13	(SEQ ID NO: 53)
538	TGYTCGTCNGCCATYTCGFI	GGFI-13	(SEQ ID NO: 54)
539	CCDATNACCATNCGNACYTT!	GGFI-17	(SEQ ID NO: 55)
540	GCHGCCUACYTGTGAC!	GGFII-1	(SEQ ID NO: 56)
541	GGYTCGGYTCCATTAUUA!	GGFII-2	(SEQ ID NO: 57)
542	CCYTCDATNACUACUACCA!	GGFII-4	(SEQ ID NO: 58)
543	TGCGCAARTANCGKCI	GGFI-11	(SEQ ID NO: 59)
544	GCHGCHAGKCYTCYTTNGC!	GGFI-14	(SEQ ID NO: 60)
545	GCHGCTAANGCYTCYTTNGC!	GGFI-14	(SEQ ID NO: 61)
546	TTYTTNGCYTCAGACUAA!	GGFI-15	(SEQ ID NO: 62)
551	TTYTTNGCYTCYAAUACAA!	GGFI-15	(SEQ ID NO: 63)
568	TCNACUAGYTCYTCAC!	GGFII-8	(SEQ ID NO: 64)
569	TCNACYAAATCYTCAC!	GGFII-8	(SEQ ID NO: 65)
609	CATRTAYTCGCHGARTGCG!	GGFII-12	(SEQ ID NO: 66)
610	CATRTAYTCGCHGRTGCG!	GGFII-12	(SEQ ID NO: 67)
649	HGARTGCGYAAHGANGCYTT!	GGFII-12	(SEQ ID NO: 68)
650	HGARTGCGUAGHGANGCYTT!	GGFII-12	(SEQ ID NO: 69)
651	ACTRTGCGYAAHGANGCYTT!	GGFII-12	(SEQ ID NO: 70)
652	ACTRTGCGUAGHGANGCYTT!	GGFII-12	(SEQ ID NO: 71)
653	HGARTGCGYAACTNGCYTT!	GGFII-12	(SEQ ID NO: 72)
654	HGARTGCGUACTNGCYTT!	GGFII-12	(SEQ ID NO: 73)
655	ACTRTGCGYAACTNGCYTT!	GGFII-12	(SEQ ID NO: 74)
656	ACTRTGCGUACTNGCYTT!	GGFII-12	(SEQ ID NO: 75)
659	ACHAGGAAATGGCTGKGA!	GGFI-13	(SEQ ID NO: 76)
660	ACHAGGAAATGGCTGKGA!	GGFI-13	(SEQ ID NO: 77)
661	CAYCARTNTGGCGGCGAA!	GGFII-1	(SEQ ID NO: 78)
662	TTYGTGCTNATGAGCGTAA!	GGFII-4	(SEQ ID NO: 79)
663	AARGCKAYGCKAYAGKGA!	GGFI-1	(SEQ ID NO: 80)
664	GARGCYTTGCKGCTTGA!	GGFI-14	(SEQ ID NO: 81)
665	GTHGCTTCGTCUAACTT!	GGFII-8	(SEQ ID NO: 82)
666	GTGCGTATGTGTCUAACTT!	GGFII-8	(SEQ ID NO: 83)
694	NACYTTTTTGAUATTTGCG!	GGFI-17	(SEQ ID NO: 87)

FIGURE 20

18/57

PCR PRIMERS FOR FACTOR I AND FACTOR II

FIGURE 22

Degenerate PCR primers

Oligo	Sequence	Peptide	
657	CCGAATTCTCCAGGARAACACACGAYCCNGGI	GGPII-17	(SEQ ID NO: 86)
658	AAGGATCTCTCAGNGTGTATGACCCATNACCATNGG!	GGPII-17	(SEQ ID NO: 87)
667	CCGAATTCTCCAGGCGAYATYCCNGARTAYATG!	GGPII-12	(SEQ ID NO: 88)
668	CCGAATTCTCCAGGCGAYATYCCNGARTAYATG!	GGPII-12	(SEQ ID NO: 89)
669	AAGGATCTCTCAGGCGAYATYCCNGARTAYATG!	GGPII-12	(SEQ ID NO: 90)
670	AAGGATCTCTCAGGCGAYATYCCNGARTAYATG!	GGPII-12	(SEQ ID NO: 91)
671	CCGAATTCTCCAGGCGAYATYCCNGARTAYATG!	GGPII-1	(SEQ ID NO: 92)
672	CCGAATTCTCCAGGCGAYATYCCNGARTAYATG!	GGPII-2	(SEQ ID NO: 93)
673	CCGAATTCTCCAGGCGAYATYCCNGARTAYATG!	GGPII-3	(SEQ ID NO: 94)
674	CCGAATTCTCCAGGCGAYATYCCNGARTAYATG!	GGPII-4	(SEQ ID NO: 95)
677	AAGGATCTCTCAGGCGAYATYCCNGARTAYATG!	GGPII-1	(SEQ ID NO: 96)
678	AAGGATCTCTCAGGCGAYATYCCNGARTAYATG!	GGPII-2	(SEQ ID NO: 97)
679	AAGGATCTCTCAGGCGAYATYCCNGARTAYATG!	GGPII-3	(SEQ ID NO: 98)
680	AAGGATCTCTCAGGCGAYATYCCNGARTAYATG!	GGPII-4	(SEQ ID NO: 99)
681	CATCTAYTCTAYTCTCTCCAGGATCTCTCCAG!	GGPII-2	(SEQ ID NO: 100)
682	CCGAATTCTCCAGGCGAYATYCCNGARTAYATG!	GGPII-1	(SEQ ID NO: 101)
683	CCGCGTAAAGCTTCTCTCCAGGATCTCTCCAG!	GGPII-14	(SEQ ID NO: 102)
684	CCGCGTAAAGCTTCTCTCCAGGATCTCTCCAG!	GGPII-14	(SEQ ID NO: 103)
685	TCTGCTAAATATGCGCGAGGATCTCTCCAG!	GGPII-1	(SEQ ID NO: 104)

Unique PCR primers for Factor II

Oligo	Sequence	Comment
711	CATCTAYTCTAYTCTCTCCAGGATCTCTCCAG!	3' RACE (SEQ ID NO: 103)
712	AAGGATCTCTCAGGCGAYATYCCNGARTAYATG!	3' RACE (SEQ ID NO: 104)
713	CCGAATTCTCCAGGCGAYATYCCNGARTAYATG!	3' RACE (SEQ ID NO: 105)
721	CATCTAYTCTAYTCTCTCCAGGATCTCTCCAG!	3' RACE (SEQ ID NO: 106)
722	AAGGATCTCTCAGGCGAYATYCCNGARTAYATG!	3' RACE; ANCHORED (SEQ ID NO: 107)
723	AAGGATCTCTCAGGCGAYATYCCNGARTAYATG!	EXONS 2 (SEQ ID NO: 108)
724	CCGAATTCTCCAGGCGAYATYCCNGARTAYATG!	EXONS 2 (SEQ ID NO: 109)
725	CATCTAYTCTAYTCTCTCCAGGATCTCTCCAG!	EXONS 3-4 (SEQ ID NO: 110)
726	ATACCCGGGCTCCAGGATCTCTCCAGGATCTCTCCAG!	(SEQ ID NO: 111)
727	AAGGATCTCTCAGGCGAYATYCCNGARTAYATG!	ANCHORED (SEQ ID NO: 112)
728	ATACCCGGGCTCCAGGATCTCTCCAGGATCTCTCCAG!	EXONS 3-4 (SEQ ID NO: 113)

19/57

Summary of contiguous GGF-II cDNA structures and sequences

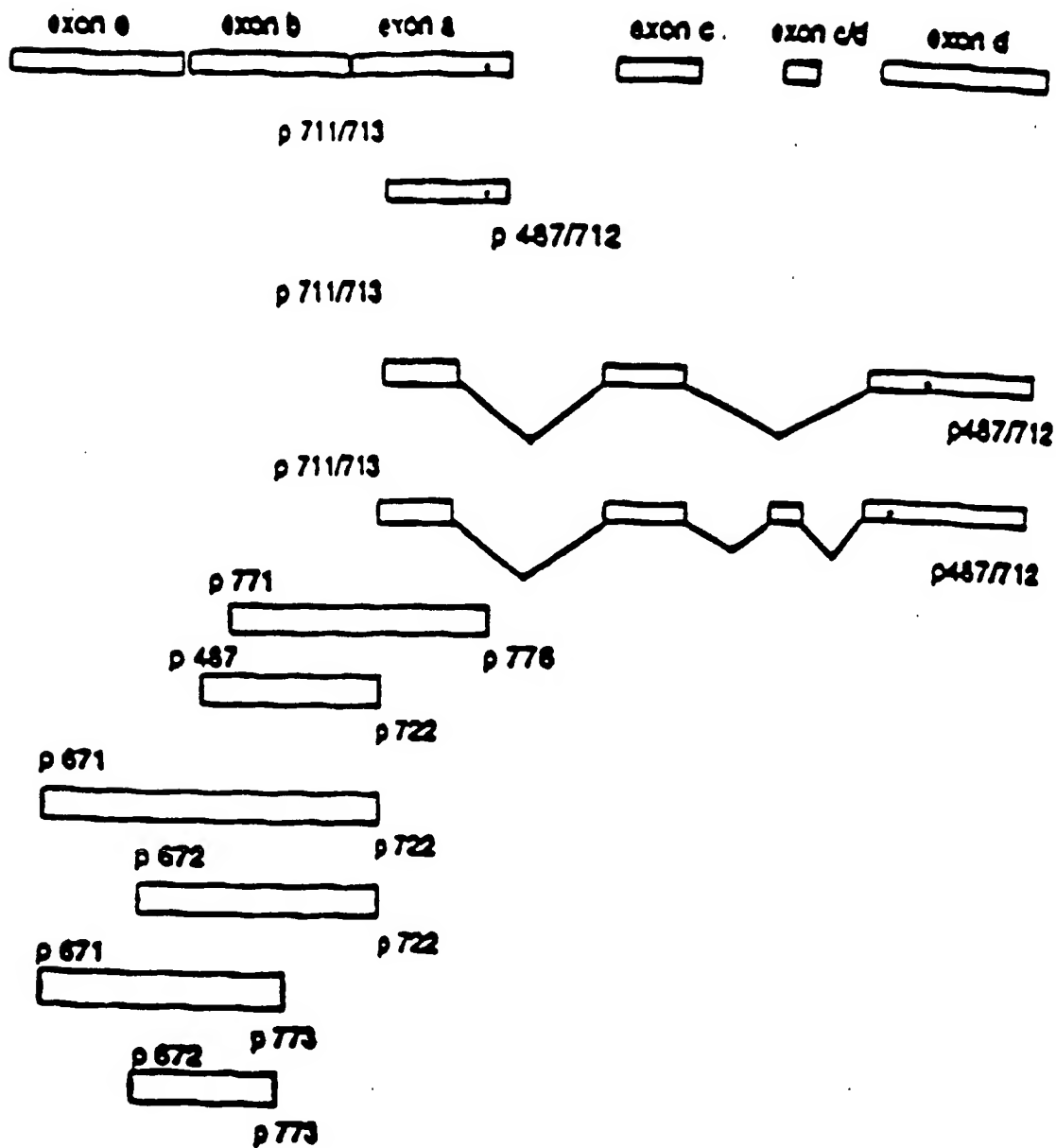


FIGURE 23

20/57

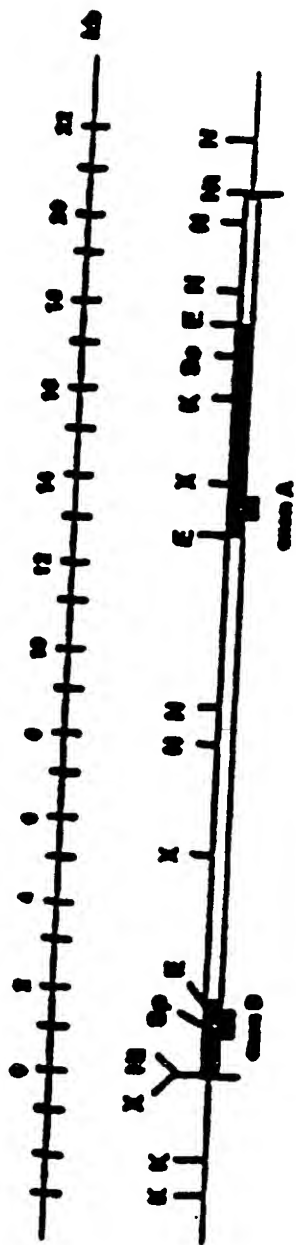


FIGURE 24

21/57

Alternative gene products of putative bovine GGF-2

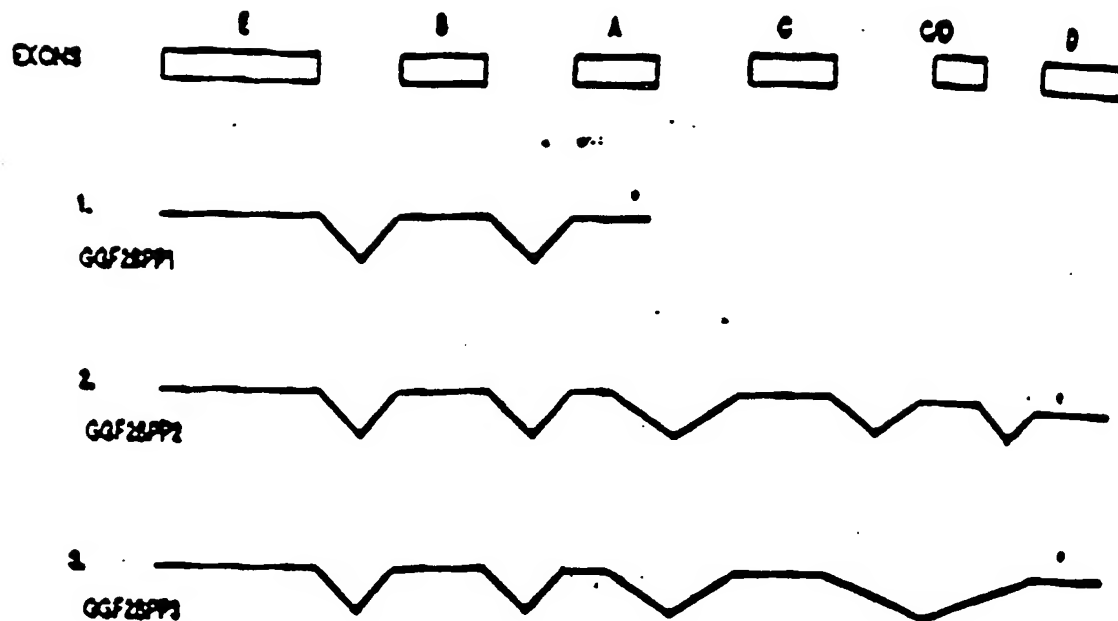


FIGURE 25

22/57

007-II polypeptides identified in deduced sequences of putative bovine 007-II proteins

Peptide	Pos.	Sequence match	
II-1	1:	VHGVAAK HGVAAK AAGLR	(SEQ ID NO: 116)
II-10	14:	DLLLV CGLV dellev ALGAV	(SEQ ID NO: 117)
II-03	21:	LGAVGPPAFPVY LLTV LGAVGHPAFPVY ALVD	(SEQ ID NO: 118) (SEQ ID NO: 119)
II-02	41:	YITTHPELVSSG KEDSR YITTHPELVSSG CPGL	(SEQ ID NO: 120) (SEQ ID NO: 121)
II-6	103:	LVL VAGSR LVL CETSS	(SEQ ID NO: 122)
I-18	112:	LYKCLKFKWTKATV CETSS eysslkfkvfkngsel SRDK	(SEQ ID NO: 123) (SEQ ID NO: 124)
II-12	151:	KSLADSGEYHCK ELRIS KSLADSGEYHCK VISKL	(SEQ ID NO: 125) (SEQ ID NO: 126)
I-07	152:	SLADEYEYHCK LRISK sladsgeyhck VISKL	(SEQ ID NO: 127) (SEQ ID NO: 128)

FIGURE 26

26/57

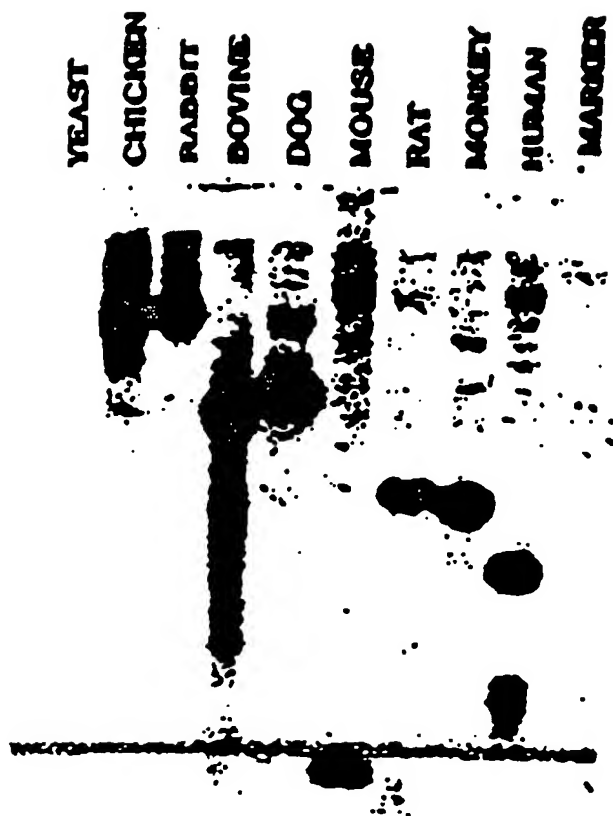


FIGURE 28

FIGURE 30 $1/c$

CODING SEGMENTS OF GLIAL GROWTH FACTOR/HEREGULIN GENE

CODING SEGMENT P1. (SEQ ID NO: 12)

[illegible]

CODING SEGMENT 2: (SEQ ID NO: 133)

[illegible]

29/57

FIGURE 30 2/8

CODING SEGMENT B: (SEQ ID NO: 134)

L P P R L K E N I S Q E S V A G S E L V
 CCTTCCCTCCCCCTTCAGAGATCAGAGTCAGGAGTCTGTCCACGTTCCAACTAG 60
 |||||
 ccttgcctccccgattgaaagagatgaaaagccagggaatcggctgcagggttccaaactag
 Q A

L R C E T S S E Y S S L K P E V P E N G
 TCCTTCGGTCCAGACCAAGTTCTCAATACTCCTCTCTCAAGTTCAAGTGGTTCAAGATG 120
 |||||
 tcccttcggctgtgaaaccagttctgaaactcctctctcagattcgaagtgggttcaagaatg
 R

S E L S R E N K P Q N I K I Q E R P G
 CGAGTCAATTAAAGCCGAAAGCAACCAACCAACCAATCAAGATACAGAAAGCCCGG 178
 |||||
 ggaatgaattgaatcgaaaaaacaaaccacaaaatatcaagatacaaaaaaagccagg
 N H K

CODING SEGMENT A: (SEQ ID NO: 135)

K S E L R I S K A S L A D S G E Y N C E
 GAAGTCAGAACTTCCCAATTAGCAAGCGTCACTCGGCTGATTCTCGAGATATATGTGCAA 60
 |||||
 gaagtcagaaactcgcattacaaagcatcactggctgattctggagagtatatgtgcaa
 H

V I S K L G N D S A S A N I I V E S N
 AGTCATCAGCAAACTAGCAATGACAGTCCCTCTGCCAACATCAACATTGTGAGTCAA 120
 |||||
 agtcatcagcaaaattaggaaatgacagtgcctctgccaaatcaccaatcgtggaaatcaa
 A

122
 ||
 123

CODING SEGMENT A': (SEQ ID NO: 136)

TCTAAACTACAGAGACTGTATTTTCATGATCATATAGTTCTGTCAATATATCTTAAAC 60

 CCTTTGCTCTCATCTTGTAGCAAGTCAGAACTTCCCAATTAGCAAGCGTCACTCGGCTG 120

S G E Y N C K V I S K L G N D S A S A N
 ATTCTCGAGCAATATATGTGCAAGTCATCAGCAAACTAGCAATGACAGTCCCTCTGCCA 180

I T I V E S N G K R C L L R A I S Q S L
 ACATCACCAATTGTGAGTCAAACGTAAGAGATGCCCTACTCGGCTCATTTCTCAGTCTC 240

R C V C K V C E N
 TAAACCAATCTCATCAAGCTATCTCTTCACTTCAATCACTCAAGCTCTCTCAATCTCAT ...

FIGURE 30 5/8

32/57

CODING SEGMENT II: (SEQ ID NO: 142)

[illegible]

L V V G I N C V V F Y C K T R E Q R R K
CTCGTGGTTGCAATGTGTGTGGTGCTACTGCMAACCAAGAACCAACCAAAg 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
cttgtggtcggcatcattgtgtgtggtggccctactgcataaaccaagaaacagcggaataag

L E D R L R Q S L R S E R N T N N N V A
CTTCATGACCGGCTTCGGCAGAGCCTTCGGTCTGAAAGAAACACCATGATGAACTAGCC 180
CTGcatgaccgtcttcggcagagccttcggctctgaacgaaacaaatgatgaacattgcc

N G P H E P N P P P E N V Q L V N Q Y V
 AACCGCCCCACCACCCCAATCCGCCCCCGAGAACCTCAGCTGGTCAATCAATACGTA
 |||||
 aatggggcctcaccatcctaaccaccccccgagaaatgtccagctgggtgaatcaataccta

TCTAATAATGTCATCTCTAGCGAGCATATTGTTGACAGAGAGCCCGCAGAGCTCTTTTTCC
tctaaaaacgtcatctccagtgagcatattggttgagagagaagcagagacatcccccttc

T S N Y T S T A E H S T V T Q T P S E
ACCACTCACTACACTTCCACAGCTCATCTCCACTACTGTCACTCAGACTCCCAGTCAC 160
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
accagtcactatacttccacagcccattcactccactactgtcacccagactccttagccac

S V S N G H T E S I I S E S E S V I V M
AGCTCGACCAATCGACACACTGAAAGCATCATTTCCGAAGCCACTGTGTCAATCCTAATC
|||
Agctggagcaacggacacactgaaagcatcctttccgaaagccaactctgtaatcgtgatg

(20)

S S V I N S R N S S P T G G P R G R L N
 TCAATCCGTAGAAAACAGTAGGCACAGCAGCCCACTGCCCCCTCAGAGGACCTCTCAAT
 |||||
 tcatccgtagaaaaacagttaggcacagcagcccaactggggggcccaagaggacgtctttaa

G L G G P R E C H S P L R H A R E T P D
 GGCTTGGAGGCCCTCGTGAATGTAAAGCTTCTCAGGCATGCCAGAGAAACCCCTGAC 340
 |||
 ggcacaggaggccctcgtgaatgtaaacagcttcctcaggcatgccagagaaacccctgat
 T

[illegible]

FIGURE 30 8/8

Human Coding Segment 2:

(SEQ ID NO: 154)

[illegible]

FIGURE 31 2/2 37/57

TCAGTTCCTCTGTCCGTGACTAGTCCGCTCTGAGCTACTCGTAGGTCGGTAAAGGCTCCAG 1380
TGTTTCTGAAATTCATCTTCAATTACTGTGATACCAATGATAGTCCCTCTCAACCACTG 1440
CAATGACAAATAAGGCTTCAAAAGCTCACTTTTATTGAGAAATAAAAATCGTTCCAG 1500
GGGACAGTCCCTCTTCTTTATAAATACCGTATCCTTCAAAAGGAGGCTGTTTAAGTTG 1560
TAACCACTACACACTTCAATGATCGTAACTTCGCTTCGGTTCAAAATGCTCTTCTTTCTG 1620
ACAAATAAACAGAAATAAAAA 1653

(SEQ ID NO: 146)

FIGURE 33 1/2

GGF23PP4 nucleotide sequence and deduced protein sequence

GAAGTCAGAACTTCCATTAGCAAGCCTCACTGGCTCAATTCTCGAGAAATATGTCCAA 60
 K S E L R I S K A S L A D S G E Y N C K
 AGTCATCAGCAACTAGCAATGACAGTGCCTCTGCCAACATCACCATTGTGGAGTCAA 120
 V I S K L G N D S A S A N I T I V E S N
 CCCCACATCCACATCTACAGCTGCCAACAGCCATCTTGTCAAGTGTCCAGAGAACGAGAA 180
 A T S T S T A G T S N L V E C A E K E E
 AACTTTCTGTGTGAATGGAGGGCACTGCTTCATGGTGAAGACCTTTCAAATCCCTCAAG 240
 T P C V N G G D C P N V K D L S N P S R
 ATACTGTGCAAGTGCCAACCTCGATTCACTGGAGCCAGATGTACTAGAAATGTGCCCCAT 300
 Y L C K C Q P G P T G A R C T E N V P N
 GAAAGTCCAAACCCAGAAAGAGCCGAGGAGCTCTACCAAGAGAGTGTCTACCAATTAC 360
 K V Q T Q E K A E E L Y Q K R V L T I T
 CGGCATTTGCAATCGCGCTGCTGCTGGTTGGCATCATGTGTGTGCTGCTCTACTGCCAAAC 420
 G I C I A L L V V G I N C V V V Y C K T
 CAAGAAACAAACGAAAGAGTTCAATGACCGGCTTCCGCAAGAGCTTCCGTCTGAAGAA 480
 K K Q R K K L N D R L R Q S L R S E R N
 CACCATCATCAACCTAGCCAACTGCCCCAGCAACCCCAATCCGCCCCCGAGAACCTGCCA 540
 T N N N V A N G P N N P N P P P E N V Q
 GCTGCTGAATCAATACGTATCTMAAAATGTTCATCTCTAGCGAGCATATTGTTGAGAGACA 600
 L V N Q Y V S K N V I S S E E I V E R E
 GCGGAGAGCTCTTTTCCACAGTCACTACCTTCCACAGCTCAATTCCTACTACTGT 660
 A E S S P S T S E Y T S T A E E S T T V
 CACTCAGACTCCCACTCAAGCTGAGCAATGGAACACCTCAAGCATCAATTCCGAAAG 720
 T Q T P S E S N S N G E T E S I I S E S
 CCACTCTGTCACTGTCATGTTCATCCCTAGAAACAGTAGGCACAGCAGCCCACTGGGG 780
 N S V I V N S S V E N S R N S S P T G G
 CCCCAGAGGAGCTCTCAATGCTTCCGAGGCCCTCTGTGAATGTAAAGCTTCTCAAGCA 840
 P R G R L N G L G G P R E C N S P L R E
 TGCCAGAGAAACCCCTCACTCTACCCAGACTCTCTCTCACTCAAGACATTAACCTTAT 900
 A R E T P D S Y R D S P E S E R E N L I
 AGCTCAGCTAAGGAGAAACAGGCCCAAGATCCAAATGCAATCAGATCCAGCTTTCCGC 960
 A E L R R N K A N R S K C N Q I Q L S A
 AACTCATCTTAGAGCTTCTTCCATTCCCAATTGGGCTTCAATTGTCTAAGACCCCTTGGCC 1020
 T N L R A S S I P N W A S P S K T P W P
 TTAAAGCAAGGTATGTATCAGCAATGACCAACCCCGGCTGTATGTACCTGTAGATTTCCA 1080
 L G R Y V S A N T T P A R N S P V D F N
 CACGCCAAGCTCTCCCAAGCTCACTCCCTTCCCAAAATGTCTCTCCCTGTCTCAGCAGCA 1140
 T P S S P K S P P S E N S P P V S S T

FIGURE 33 2/2

40/57

GGTCTCCATGCCCTCCATGCGCGTCAGTCCCTTCTGCGAAGACGAGAGACACCCCTGCTCTT 1200
V S N P S N A V S P P V E E E R P L L L
TCTACGCCACCAACCGCTGCGGAGAGTATGACCACCAACCGCCAGCAATTCACTCTT 1260
V T P P R L R E I Y D N N A Q Q P N S P
CCTCTCAACCCCGCCATGAGAGCCACAGCCTGCCCCCAGCCCCCTGAGGATAGTCCA 1320
H C N P A H E S N S L P P S P L R I V E
CGATCAGGAATATCAAAAGGACCCAGGAGTACCAACCAAGCTCAAGAGCCCGTTAAGAACT 1380
D E E Y E T T Q E Y E P A Q E P V K E L
CACCAACAGCAGCCCGCGCCCAAAAGAACCAAGCCCAATGGTCAATTGCCCAAGGTT 1440
T N S S R R A K R T K P N G E I A E R L
CGAAATCGACAAACACACAGCGCGTGAACAGCACTAACTCAAGAGCCCAACAGAGGATGA 1500
E N D N N T G A D S S N S E S E T E D E
AAGAGTAGGAGAGATACGGCCTTCTGCGCCATACAGAACCCCGCTGCGCAGCCAGTCTCCA 1560
R V G E D T P P L A I Q N P L A A S L E
CGCGCGCCCTGCGCTTCCGCTGCTGCAACAGGACTAAACCCAAACAGCGCGCTTCTCTCC 1620
A A P A P R L V D S R T N P T G G P S P
GCAGGAACAAATTGCAAGCCACGGCTCTCCGTTGTAATCGCTAACCAAGACCCATACCGCTT 1680
Q E E L Q A R L S G V I A N Q D P I A V
CTAAACCCAAATACACCCATAGATTCACTGTAAACTTTATTTTATATTAATAAGTAT 1740
TCCACCTTAATTAAACAAAA 1764

FIGURE 34

GGF2¹KKCAEKEKTCVNGGECFMYKDLSNPSRYLCKCPNEFTGDRCONYVMASFY¹
GGF2²KKCAEKEKTCVNGGDCFMYKDLSNPSRYLCKCQPGFTGARCTENYPMKYQ²
DEGE ECLRKYKDFCIH - GECKYVKELRAPS — CKCQOEYFGERCGEKSNTHS³

¹(SEQ ID NO: 147)

²(SEQ ID NO: 148)

³(SEQ ID NO: 149)

FIGURE 35

200 kDa tyrosine phosphorylation compared with mitogenic activity

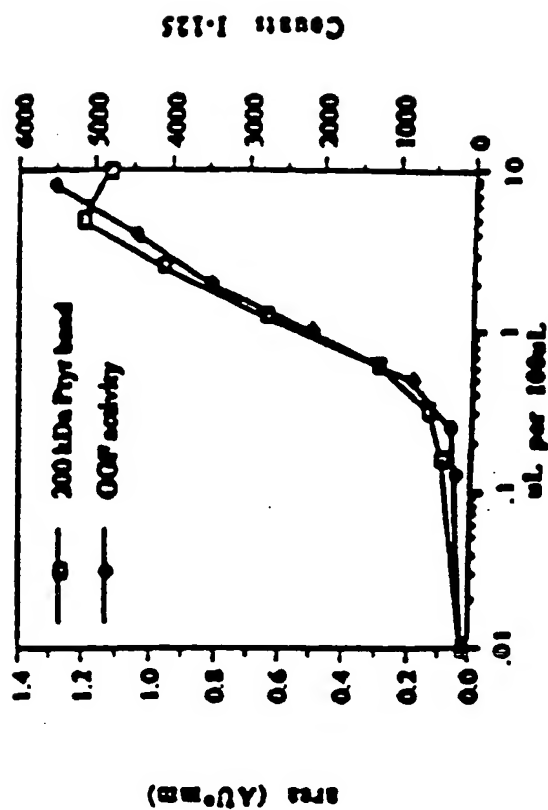


FIGURE 36 *1/2*

CG7/REGULON SPLICING VARIANTS

7-1-10

F-B-A-C-C/D-D
F-B-A-C-C/D-E
F-B-A-C-C/D-E-L
F-B-A-C-C/D-E-K-L
F-B-A-C-C/D-D'-E
F-B-A-C-C/D-D'-E-L
F-B-A-C-C/D-D'-E-K-L
F-B-A-C-C/D'-D
F-B-A-C-C/D'-E
F-B-A-C-C/D'-E-L
F-B-A-C-C/D'-E-K-L
F-B-A-C-C/D'-D'-E
F-B-A-C-C/D-D'-E-L
F-B-A-C-C/D'-D'-E-K-L
F-B-A-C-C/D-C/D'-D
F-B-A-C-C/D-C/D'-E
F-B-A-C-C/D-C/D'-E-L
F-B-A-C-C/D-C/D'-E-K-L
F-B-A-C-C/D-C/D'-D'-E
F-B-A-C-C/D-C/D'-D'-E-L
F-B-A-C-C/D-C/D'-D'-E-K-L

F-B-A-C-C-C/D-D
F-B-A-C-C-C/D-E
F-B-A-C-C-C/D-E-L
F-B-A-C-C-C/D-E-K-L
F-B-A-C-C-C/D-D'-E
F-B-A-C-C-C/D-D'-E-L
F-B-A-C-C-C/D-D'-E-K-L
F-B-A-C-C-C/D-D'-E
F-B-A-C-C-C/D-D'-E-L
F-B-A-C-C-C/D-D'-E-K-L
F-B-A-C-C-C/D-D'-E
F-B-A-C-C-C/D-D'-E-L
F-B-A-C-C-C/D-D'-E-K-L
F-B-A-C-C-C/D-D'-E
F-B-A-C-C-C/D-D'-E-L
F-B-A-C-C-C/D-D'-E-K-L
F-B-A-C-C-C/D-D'-E
F-B-A-C-C-C/D-D'-E-L
F-B-A-C-C-C/D-D'-E-K-L

7-8-41'

F-E-B-A-C-C/D-B
 F-E-B-A-C-C/D-E
 F-E-B-A-C-C/D-E-L
 F-E-B-A-C-C/D-E-K-L
 F-E-B-A-C-C/D-D'-E
 F-E-B-A-C-C/D-D'-E-L
 F-E-B-A-C-C/D-D'-E-K-L
 F-E-B-A-C-C/D'-B
 F-E-B-A-C-C/D'-E
 F-E-B-A-C-C/D'-E-L
 F-E-B-A-C-C/D'-E-K-L
 F-E-B-A-C-C/D'-D'-E
 F-E-B-A-C-C/D'-D'-E-L
 F-E-B-A-C-C/D'-D'-E-K-L
 F-E-B-A-C-C/D-C/D'-B
 F-E-B-A-C-C/D-C/D'-E
 F-E-B-A-C-C/D-C/D'-E-L
 F-E-B-A-C-C/D-C/D'-E-K-L
 F-E-B-A-C-C/D-C/D'-D'-E
 F-E-B-A-C-C/D-C/D'-D'-E-L
 F-E-B-A-C-C/D-C/D'-D'-E-K-L

~~F-E-B-A-C-C-C/D-D~~
~~F-E-B-A-C-C-C/D-E~~
~~F-E-B-A-C-C-C/D-E-L~~
~~F-E-B-A-C-C-C/D-E-K-L~~
~~F-E-B-A-C-C-C/D-D'-E~~
~~F-E-B-A-C-C-C/D-D'-E-L~~
~~F-E-B-A-C-C-C/D-D'-E-K-L~~
~~F-E-B-A-C-C-C/D'-D~~
~~F-E-B-A-C-C-C/D'-E~~
~~F-E-B-A-C-C-C/D'-E-L~~
~~F-E-B-A-C-C-C/D'-E-K-L~~
~~F-E-B-A-C-C-C/D'-D'-E~~
~~F-E-B-A-C-C-C/D'-D'-E-L~~
~~F-E-B-A-C-C-C/D'-D'-E-K-L~~
~~F-E-B-A-C-C-C/D-C/D'-D~~
~~F-E-B-A-C-C-C/D-C/D'-E~~
~~F-E-B-A-C-C-C/D-C/D'-E-L~~
~~F-E-B-A-C-C-C/D-C/D'-E-K-L~~
~~F-E-B-A-C-C-C/D-C/D'-D'-E~~
~~F-E-B-A-C-C-C/D-C/D'-D'-E-L~~
~~F-E-B-A-C-C-C/D-C/D'-D'-E-K-L~~

FIGURE 36 2/2

007/HEBECULIN SPLICING VARIANTS CONTINUED

E-B-A'

E-B-A-C-C/D-D
 E-B-A-C-C/D-E
 E-B-A-C-C/D-E-L
 E-B-A-C-C/D-E-K-L
 E-B-A-C-C/D-D'-E
 E-B-A-C-C/D-D'-E-L
 E-B-A-C-C/D-D'-E-K-L
 E-B-A-C-C/D'-D
 E-B-A-C-C/D'-E
 E-B-A-C-C/D'-E-L
 E-B-A-C-C/D'-E-K-L
 E-B-A-C-C/D'-D'-E
 E-B-A-C-C/D'-D'-E-L
 E-B-A-C-C/D'-D'-E-K-L
 E-B-A-C-C/D-C/D'-D
 E-B-A-C-C/D-C/D'-E
 E-B-A-C-C/D-C/D'-E-L
 E-B-A-C-C/D-C/D'-E-K-L
 E-B-A-C-C/D-C/D'-D'-E
 E-B-A-C-C/D-C/D'-D'-E-L
 E-B-A-C-C/D-C/D'-D'-E-K-L

E-B-A-C-C-C/D-D
 E-B-A-C-C-C/D-E
 E-B-A-C-C-C/D-E-L
 E-B-A-C-C-C/D-E-K-L
 E-B-A-C-C-C/D-D'-E
 E-B-A-C-C-C/D-D'-E-L
 E-B-A-C-C-C/D-D'-E-K-L
 E-B-A-C-C-C/D'-D
 E-B-A-C-C-C/D'-E
 E-B-A-C-C-C/D'-E-L
 E-B-A-C-C-C/D'-E-K-L
 E-B-A-C-C-C/D'-D'-E
 E-B-A-C-C-C/D'-D'-E-L
 E-B-A-C-C-C/D'-D'-E-K-L
 E-B-A-C-C-C/D-C/D'-D
 E-B-A-C-C-C/D-C/D'-E
 E-B-A-C-C-C/D-C/D'-E-L
 E-B-A-C-C-C/D-C/D'-E-K-L
 E-B-A-C-C-C/D-C/D'-D'-E
 E-B-A-C-C-C/D-C/D'-D'-E-L
 E-B-A-C-C-C/D-C/D'-D'-E-K-L

EGFL1

45/57

AGCCATCTTGTCAAGTGTGCAGAGACGAGAAAAGCTTTCTGTGTCAATGGAGCCGACTGC
S H L V K C A E K E K T P C V N G G E C
TTCATGGTGAAAGACCTTTCAAATCCCTCAAGATACTTGTGCAAGTGGCCAAATGAGTTT
P M V K D L S N P S R Y L C K C P N E P
ACTGGTGATCGCTGCCAAAAGTACGTAATGGCCAGCTTCTACAGTACGTCCTCACTCCCTTT
T G D R C Q N Y V N A S P Y S T S T P P
CTGTCTCTGCTCAATAG
L S L P E •

(SEQ ID NO: 150)

FIGURE 37

EGFL2

46/57

AGCCATCTTGTCAAGTGTGCAGAGAGGAGAAAACCTTTCTGTGTGAATGGAGGCGAGTGC
S H L V K C A E E E K T P C V N G G E C
TTCATCGTGAAAGACCTTTCAAATCCCTCAAGATACTTGTGCAAGTCCCAACCTCGATTG
P N V K D L S N P S R Y L C K C Q P G P
ACTCGAGCGAGATGTACTGAGAAATGTGCCCATGAAAGTCCAAACCCAGAAAAGCGGAG
T G A R C T E N V P N K V Q T Q E K A E
GAGCTCTACTAA
E L Y *

(SEQ ID NO: 15/)

FIGURE 38

EGFLJ

47/57

AGCCATCTTGTCAAGTGTGCAAGAGAGGAGAAAAGTTTCTGTGTGAATGGAGGCGAGTGC
S E L V R C A E K E K T P C V N G G E C
TTCATCGTGAAAGACCTTTCAATCCCTCAAGATACTGTGCAAGTCCCAATGAGTTT
P H V R D L S N P S R Y L C K C P H E P
ACTCGTGATCCCTGCCAAACTACCTAATGGCCAGCTTCTACAAAGCCGAGGAGCTCTAC
T G D R C Q N Y V H A S P Y K A E E L Y
TAA
•

(SEQ ID NO: 132)

FIGURE 39

EGFLA

AGCCATCTTGTCAAGTGTGCAAGAGAGGAGAAAACTTTCTGTGTCAATCGAGGCGAGTGC
S H L V K C A E E E K T P C V N G G E C
TTCATCGGTGAAGACCTTTCAATCCCTCAAGATACTTGTGCAAGTGCCCAATCAGTTT
P M V K D L S M P S R Y L C E C P N E P
ACTCGGTGATCGGTGCCAAACTACCTAATCGCCAGCTTCTACAGCATCTTGGGATTGAA
T G D R C Q N Y V M A S P Y E N L G I E
TTTATCGAGAAAGCGGAGGAGCTCTACTAA
P M E K A E E L Y •

(SEQ ID NO: 153)

FIGURE 40

EGFLS

49/57

AGCCATCTTGTCTAGTGTGCAGAGAGGAGAAAAGTTTCTGTGTGAATGGAGCCGAGTGC
S H L V K C A E K E K T P C V H G G E C
TTCACTGGTGAAGACCTTTCAATCCCTCAAGATACTTGTCTAGTCCCACTTGGATTG
P H V K D L S H P S R Y L C K C Q P G P
ACTCGAGCGAGATGTACTGAGAAATGTGCCCATGAAGTCCAAACCCAGAAAAGTCCCA
T G A R C T E N V P H K V Q T Q E K C P
AATGAGTTTACTGGTGATCGCTGCCAAAAGTACCTAATGCCAGCTTCTACAGTACGTCC
N E P T G D R C Q H Y V N A S P Y S T S
ACTCCCTTTCTGTCTCTGCTGAAATAG
T P P L S L P E .

(SEQ ID NO: 134)

FIGURE 41

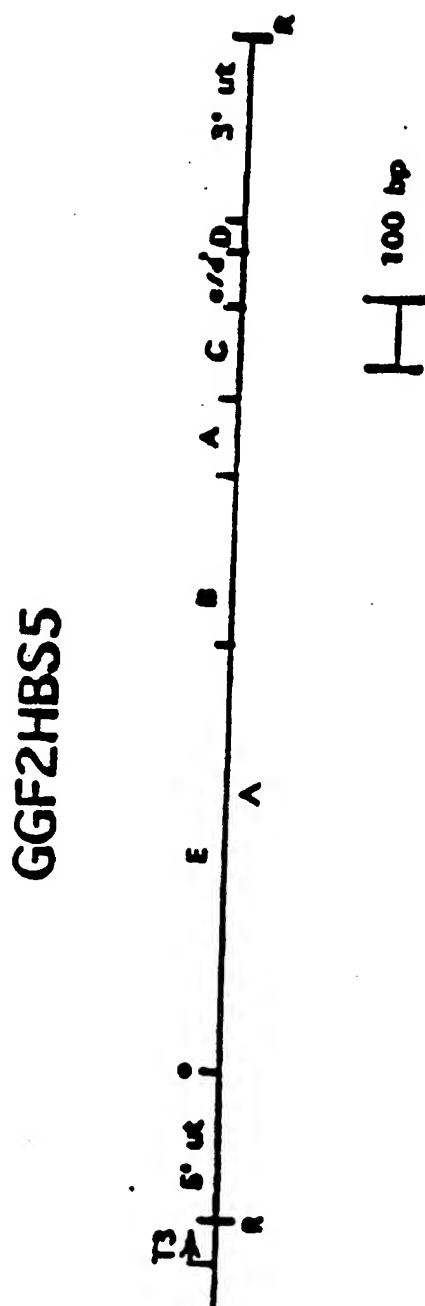
EGFL6

50/57

AGCCATCTTGTCAAGTGTGCAGAGAGGAGAAAACTTTCTGTGTGAATGGAGGCCAGTGC
S H L V K C A E K E K T P C V H G G E C
TTCAATGGTGAAGACCTTTCAATCCCTCAAGATACTTGTGCAGTGCCAACCTGGATTG
P H V K D L S H P S R Y L C K C Q P G F
ACTCGAGCGAGATGTACTGAGAAATGTGCCCATGAAGTCCAAACCCAGAAAAGTCCCCA
T G A R C T E H V P H K V Q T Q E K C P
AATGAGTTTACTGGTGATCGCTGCCAAACTACGTAATGCCAGCTTCTACAAAGCCGAG
H E F T G D R C Q H Y V H A S F Y K A E
GAGCTCTACTAA
E L Y •

(SEQ ID NO: 135)

FIGURE 42

FIGURE 43

53/57

<p> GGGGGGGGGGGAGGAGGGGGGGCTATCTGGTTGAGGCTGACGAGGTGTGGGGGGTGAAGGGC E P G E E A P Y L V E V E G V W A A E A GGVII-1 & GGVII-11 </p> <p> GGGGGGTGAAGGAGGAGTGGCTGCTGCTGACGGTGGGGGGTGGGGAGCTGGGGGGGAGGGGGGGC G G L E E D S L L T V R L G T W G E P A D L L L I V L G A W G P P A GGVII-10 GGVII-3 </p> <p> TTGGGCTGCTGGGGAGGGCTGAGGAGGAGGAGGAGGAGTACATCTGCTGCTGAGTGAAGGGGAGC P P S C G E L E E D S E Y I P P M E P D P P V E Y Y I P P M E P E GGVII-2 </p> <p> GGGAGGAGGAGGAGGGGGGGGGGGGGGGGGGGGGTGGAGGGCTGCTTTGGGGGGCTCTGGAGAGG A M S T S R A P A A P E A S P P P L E T A I S S G </p> <p> GGGGGGAGCTCTAGAGAGAGGCTGAGGGGGGTGCTGTGTGAGGGGCTGGGGCTTGGCTGGC G R N L E E E V S E V L C E R C A L P P </p> <p> CGATTGAAGGAGATGAAGGGAGGAGATGGCTGCTGAGGCTTCCAGGAGTGTCTCTGGGTGT G L E E M E S G E S A A G S E L V L E C L V L R GGVII-4 </p> <p> GAGGAGGAGTGTGAATGCTGCTGCTGCTGAGTGAAGGTGTGAAGATGGAGTGAATG E T S S E Y S S L R P E M P E M G E L </p> <p> AGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT E E E E P G H I K I G E E P G E S E L </p> <p> GGATTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG E I N E A S L A D S G E Y M C E V I S E E A S L A D S G E Y M E E GGVII-12 </p> <p> TTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG L G E D S A S A N I T I V E S E A T S T </p>	<p>780</p> <p>840</p> <p>900</p> <p>960</p> <p>1020</p> <p>1080</p> <p>1140</p> <p>1200</p> <p>1260</p> <p>1320</p>
---	---

FIGURE 44 2/3

54/57

TOCLOCTOOCALMOCCICTOTUAAITOTOCOCALMOCCALMACTTCTOTOTB	1380	•
S T T O T S E L V E C A E E E E T P C V		•
AAOCALMOCCIMOTCTTCAITOTUAAIACTTTCALMOCCCTOOCALMACTTCTOTOCALMO	1440	•
N O O E C P M V E D L S E P S E Y L C E		•
TOOCALMACTOITTACTOTUATOCCTOOCALMACTOCTALTOOCALMACTTCTACACT	1500	•
C P M E P T O D E C Q N Y V N A S P Y S		•
AOCTCALCTCTCTTCTOTCTCTOCTUAAITOCLOCAITCTCTAOITTOCTCTCTCTCT	1560	•
T S T P P L S L P E		•
TOTTOCTOCLCTCTOCTCIGITTCALCTOCTALMACTALITOTCTCTACCLATCTUAT	1620	•
TTACTOOCCTCTOOCCTOOCALMAACITTAACALMAOCCUATTOTATTACTTCTCTO	1680	•
TTOOCALCTOITTOOCCTCTCIGITTAUAAITOTCTOTUAAOCTTOOCALTOITTCTOOCAL	1740	•
TTUATATTAUATOTUATCUMATTCATATOTCALATCTMOCCOTUAAITATUATAS	1800	•
AAOCCATTTCUATOTCTCITTTATTAUAAITUAAITATCTCTUAACTOCT	1860	•
ATCTTCTTAUAAITAOCCICATCTOCTAIAAOCTTOTCTCTUOCTUAAOCCATATOC	1920	•
ACTTAAITCAITOTTACTTAITTTUATCTCIGATOTOTTAITTOTCICUMUAACT	1980	•
AAUAAOIAUUUUUUUU	2000	•

FIGURE 44 3/3

FIGURE 45B

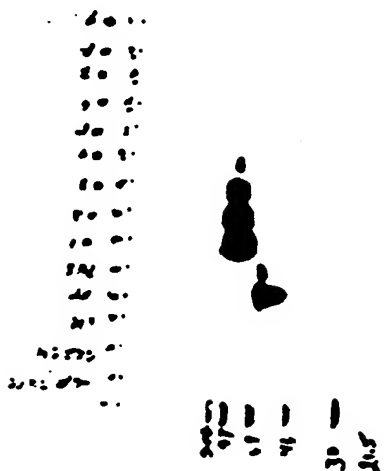
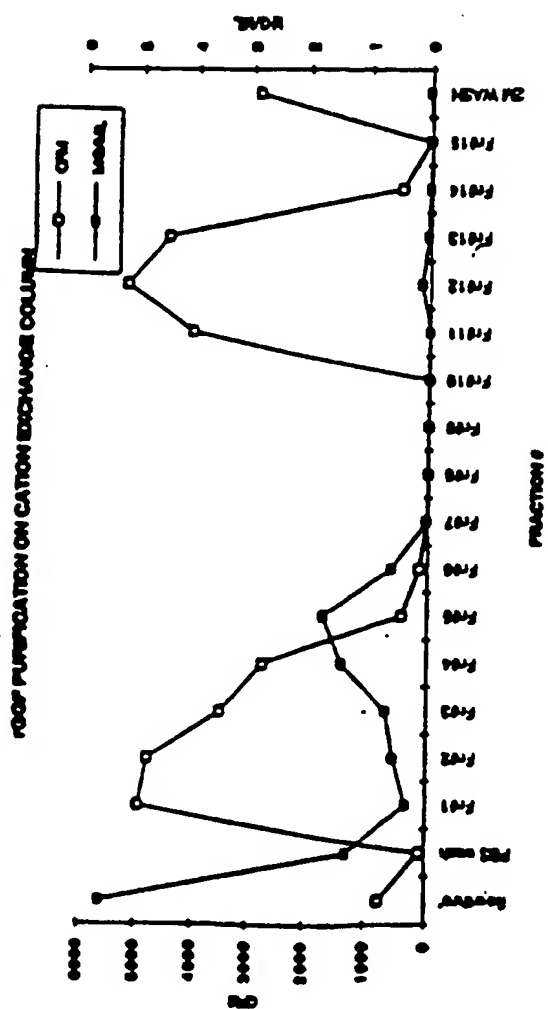


FIGURE 45A



[illegible]

FIGURE 46

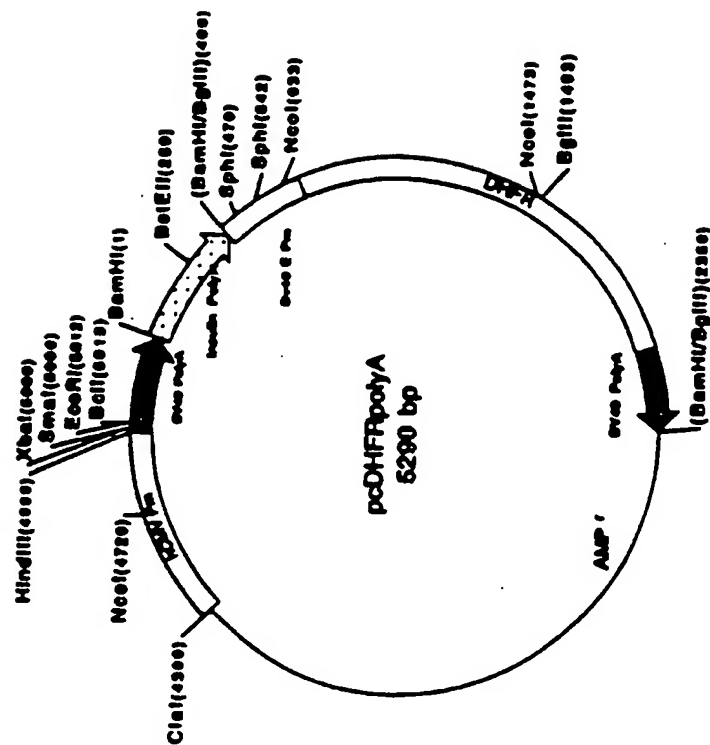


FIGURE 47

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US94/05083

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :A61K 37/10, 37/36

US CL :514/8, 12; 530/399

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : N/A

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	N/A	



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*G*	document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

22 SEPTEMBER 1994

Date of mailing of the international search report

30 SEP 1994

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Authorized officer

SHELLY GUEST CERMAK

Facsimile No. (703) 305-3230

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/05083

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-91
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

because applicants have failed to submit a searchable computer Sequence Listing, and each of the claims encompasses DNA or amino acid sequences.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.